GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:44:31 ; Search time 29.025 Seconds
(without alignments)
273.726 Million cell updates/sec

Title: US-10-074-620-1
Perfect score: 18
Sequence: 1 ggctggtgtcacctgtta 18

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16	88.9	20	1	US-08-197-791-28	Sequence 28, Appl
c	2	16	88.9	50	1	US-08-171-389-546	Sequence 546, App
c	3	16	88.9	50	1	US-08-123-936-546	Sequence 546, App
c	4	16	88.9	50	2	US-08-475-228A-546	Sequence 546, App
c	5	16	88.9	50	3	US-08-482-080A-546	Sequence 546, App
c	6	16	88.9	50	4	US-09-354-947-546	Sequence 546, App
c	7	16	88.9	50	5	PCT-US93-12388-546	Sequence 546, App
c	8	16	88.9	3833	1	US-08-917-320-18	Sequence 18, Appl
c	9	16	88.9	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
c	10	16	88.9	5931	3	US-08-783-774-1	Sequence 1, Appli
c	11	16	88.9	5931	4	US-09-556-706B-1	Sequence 1, Appli
	12	15	83.3	1808	4	US-09-171-710-5	Sequence 5, Appli
	13	15	83.3	3065	4	US-09-171-710-3	Sequence 3, Appli
	14	14	77.8	5467	1	US-07-745-206A-12	Sequence 12, Appl
	15	14	77.8	5467	2	US-08-311-363-12	Sequence 12, Appl
c	16	13	72.2	261	4	US-09-657-453A-24	Sequence 24, Appl
c	17	13	72.2	308	4	US-09-141-027-9	Sequence 9, Appli
c	18	13	72.2	308	4	US-09-617-804-9	Sequence 9, Appli
c	19	13	72.2	361	4	US-09-702-705-1159	Sequence 1159, Ap
c	20	13	72.2	361	4	US-09-736-457-1159	Sequence 1159, Ap
	21	13	72.2	504	4	US-09-228-986-40	Sequence 40, Appl
	22	13	72.2	794	4	US-09-173-300-10	Sequence 10, Appl
c	23	13	72.2	1124	3	US-09-221-456-1	Sequence 1, Appli
c	24	13	72.2	1124	4	US-09-558-740-1	Sequence 1, Appli
c	25	13	72.2	1302	4	US-09-328-352-3822	Sequence 3822, Ap
c	26	13	72.2	1716	4	US-09-328-352-2907	Sequence 2907, Ap
	27	13	72.2	1742	3	US-09-099-676-2	Sequence 2, Appli
	28	13	72.2	1742	3	US-09-565-910-2	Sequence 2, Appli
c	29	13	72.2	2972	1	US-08-454-455-3	Sequence 3, Appli
c	30	13	72.2	3789	1	US-08-454-455-5	Sequence 5, Appli
	31	13	72.2	3959	1	US-08-474-067-1	Sequence 1, Appli
	32	13	72.2	3959	2	US-08-474-068A-1	Sequence 1, Appli
	33	13	72.2	3959	2	US-08-472-481-1	Sequence 1, Appli
c	34	13	72.2	5137	5	PCT-US96-01314-39	Sequence 39, Appl
c	35	13	72.2	5138	2	US-08-476-062A-39	Sequence 39, Appl
	36	13	72.2	36941	4	US-08-311-731A-130	Sequence 130, App
c	37	13	72.2	51552	4	US-09-733-294A-30	Sequence 30, Appl
	38	13	72.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	39	13	72.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	40	12	66.7	54	4	US-08-584-040-8221	Sequence 8221, Ap
c	41	12	66.7	54	4	US-08-679-645-616	Sequence 616, App
c	42	12	66.7	54	4	US-09-371-772B-11131	Sequence 11131, A
	43	12	66.7	60	4	US-09-128-354-12	Sequence 12, Appl
	44	12	66.7	60	4	US-09-128-354-19	Sequence 19, Appl
c	45	12	66.7	275	4	US-09-313-294A-3690	Sequence 3690, Ap
c	46	12	66.7	285	4	US-09-313-294A-1993	Sequence 1993, Ap
c	47	12	66.7	289	4	US-09-313-294A-1541	Sequence 1541, Ap
	48	12	66.7	333	4	US-09-702-705-384	Sequence 384, App
	49	12	66.7	333	4	US-09-736-457-384	Sequence 384, App
	50	12	66.7	382	4	US-09-702-705-240	Sequence 240, App
	51	12	66.7	382	4	US-09-736-457-240	Sequence 240, App
c	52	12	66.7	633	3	US-09-328-111-48	Sequence 48, Appl
	53	12	66.7	693	4	US-09-629-645A-10	Sequence 10, Appl
	54	12	66.7	727	4	US-09-629-645A-21	Sequence 21, Appl
	55	12	66.7	771	3	US-08-777-708C-7	Sequence 7, Appli
c	56	12	66.7	866	4	US-09-620-312D-682	Sequence 682, App
	57	12	66.7	984	4	US-09-252-991A-11081	Sequence 11081, A
c	58	12	66.7	1018	4	US-08-976-259-93	Sequence 93, Appl
c	59	12	66.7	1152	2	US-09-036-582-34	Sequence 34, Appl
	60	12	66.7	1202	4	US-09-620-312D-97	Sequence 97, Appl
c	61	12	66.7	1314	4	US-09-252-991A-13338	Sequence 13338, A
	62	12	66.7	1359	4	US-09-252-991A-12612	Sequence 12612, A
c	63	12	66.7	1585	4	US-09-071-035-43	Sequence 43, Appl
c	64	12	66.7	1659	4	US-09-071-035-41	Sequence 41, Appl
	65	12	66.7	1722	4	US-09-489-847-93	Sequence 93, Appl
	66	12	66.7	1953	2	US-08-557-122A-1	Sequence 1, Appli
	67	12	66.7	1953	4	US-09-262-666-1	Sequence 1, Appli
	68	12	66.7	2014	4	US-09-620-312D-1037	Sequence 1037, Ap
	69	12	66.7	2169	4	US-09-617-145-1	Sequence 1, Appli
	70	12	66.7	2337	4	US-09-252-991A-11020	Sequence 11020, A
	71	12	66.7	3035	4	US-09-620-312D-286	Sequence 286, App
	72	12	66.7	3188	3	US-08-943-731-183	Sequence 183, App
	73	12	66.7	3197	4	US-09-620-312D-285	Sequence 285, App
	74	12	66.7	3855	3	US-08-974-549A-4	Sequence 4, Appli
	75	12	66.7	3855	4	US-08-912-951-4	Sequence 4, Appli
	76	12	66.7	4015	3	US-08-851-843A-224	Sequence 224, App
	77	12	66.7	4015	3	US-08-974-549A-1	Sequence 1, Appli
	78	12	66.7	4015	3	US-08-854-050-224	Sequence 224, App
	79	12	66.7	4015	4	US-09-430-323-224	Sequence 224, App
	80	12	66.7	4015	4	US-09-572-423B-3	Sequence 3, Appli
	81	12	66.7	4015	4	US-09-128-354-1	Sequence 1, Appli
	82	12	66.7	4015	4	US-09-675-321-1	Sequence 1, Appli
	83	12	66.7	4015	4	US-09-052-919-1	Sequence 1, Appli
	84	12	66.7	4015	4	US-08-912-951-1	Sequence 1, Appli
	85	12	66.7	4015	4	US-09-733-294A-3	Sequence 3, Appli
	86	12	66.7	4029	3	US-08-851-843A-173	Sequence 173, App
	87	12	66.7	4029	3	US-08-974-549A-292	Sequence 292, App
	88	12	66.7	4029	3	US-08-854-050-173	Sequence 173, App
	89	12	66.7	4029	4	US-09-430-323-173	Sequence 173, App
	90	12	66.7	4037	3	US-08-974-549A-343	Sequence 343, App
	91	12	66.7	4164	1	US-08-188-582-3	Sequence 3, Appli
	92	12	66.7	4164	1	US-08-646-715-3	Sequence 3, Appli
	93	12	66.7	4200	4	US-08-912-951-6	Sequence 6, Appli
	94	12	66.7	4335	3	US-08-974-549A-6	Sequence 6, Appli
	95	12	66.7	6911	1	US-08-311-174-4	Sequence 4, Appli
c	96	12	66.7	11770	4	US-08-961-527-172	Sequence 172, App
c	97	12	66.7	14654	4	US-08-961-527-106	Sequence 106, App
	98	12	66.7	15418	4	US-09-783-203-1	Sequence 1, Appli
	99	12	66.7	18318	1	US-08-414-926A-6	Sequence 6, Appli
	100	12	66.7	18318	2	US-08-926-922-6	Sequence 6, Appli

101	12	66.7	18318	3	US-09-253-682-6	Sequence 6, Appli
102	12	66.7	18318	3	US-09-527-657-6	Sequence 6, Appli
c 103	12	66.7	18994	1	US-08-459-586-4	Sequence 4, Appli
c 104	12	66.7	18994	2	US-08-282-696-4	Sequence 4, Appli
105	12	66.7	20084	3	US-08-943-731-5	Sequence 5, Appli
c 106	12	66.7	29629	4	US-09-729-995-3	Sequence 3, Appli
c 107	12	66.7	30310	4	US-09-657-346A-96	Sequence 96, Appl
108	12	66.7	51552	4	US-09-733-294A-30	Sequence 30, Appl
109	12	66.7	62804	4	US-09-800-960-3	Sequence 3, Appli
110	12	66.7	169998	4	US-09-676-610B-24	Sequence 24, Appl
c 111	12	66.7	174493	4	US-09-804-471A-3	Sequence 3, Appli
112	12	66.7	197496	4	US-09-877-177A-10	Sequence 10, Appl
113	12	66.7	202001	4	US-09-734-674-3	Sequence 3, Appli
c 114	12	66.7	319608	4	US-09-539-333D-1	Sequence 1, Appli
c 115	12	66.7	319608	4	US-09-679-409-1	Sequence 1, Appli
c 116	12	66.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 117	12	66.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 118	11	61.1	20	2	US-08-450-905B-125	Sequence 125, App
c 119	11	61.1	20	3	US-07-982-759F-125	Sequence 125, App
120	11	61.1	20	4	US-09-422-978-5580	Sequence 5580, Ap

ALIGNMENTS

RESULT 1
US-08-197-791-28
; Sequence 28, Application US/08197791
; Patent No. 5556772
; GENERAL INFORMATION:
; APPLICANT: Sorge, Joseph A.
; APPLICANT: Mullinax, Rebecca L.
; TITLE OF INVENTION: NOVEL POLYMERASE COMPOSITIONS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Limbach and Limbach
; STREET: 2001 Ferry Building
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/197,791
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/164,290
; FILING DATE: 08-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bortner, Scott R.
; REGISTRATION NUMBER: 34,298
; REFERENCE/DOCKET NUMBER: STRG 20270 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-433-4150
; TELEFAX: 415-433-8716
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-197-791-28

Query Match 88.9%; Score 16; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
Db 1 GGCTGGTGTCACCTGT 16

RESULT 2
US-08-171-389-546/c
; Sequence 546, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 546:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Epstein Barr virus L2 (start site
; INDIVIDUAL ISOLATE: 90021)
US-08-171-389-546

Query Match 88.9%; Score 16; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 3
US-08-123-936-546/c
; Sequence 546, Application US/08123936
; Patent No. 5726014
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Cynthia A.
;    APPLICANT:  Cantor, Charles R.
;    APPLICANT:  Andrews, Beth M.
;    APPLICANT:  Turin, Lisa M.
;    TITLE OF INVENTION:  Screening Assay for the Detection of
;    TITLE OF INVENTION:  DNA-Binding Molecules
;    NUMBER OF SEQUENCES:  640
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genelabs Technologies, Inc.
;      STREET:  505 Penobscot Drive
;      CITY:  Redwood City
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94063
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,936
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/996,783
;      FILING DATE:  23-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/723,618
;      FILING DATE:  27-JUN-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  4600-0075.32/G19P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  546:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Epstein Barr virus L2  (start site
;      INDIVIDUAL ISOLATE:  90021)
US-08-123-936-546

  Query Match           88.9%;  Score 16;  DB 1;  Length 50;
  Best Local Similarity  100.0%;  Pred. No. 1.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 4
US-08-475-228A-546/c
; Sequence 546, Application US/08475228A
; Patent No. 5869241
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Cynthia A.
;    APPLICANT:  Cantor, Charles R.
;    APPLICANT:  Andrews, Beth M.
;    APPLICANT:  Turin, Lisa M.
;    APPLICANT:  Fry, Kirk E.
;    TITLE OF INVENTION:  Sequence-Directed DNA Binding
;    TITLE OF INVENTION:  Molecules, Compositions and Methods
;    NUMBER OF SEQUENCES:  664
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genelabs Technologies, Inc.
;      STREET:  505 Penobscot Drive
;      CITY:  Redwood City
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94063
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/475,228A
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/123,936
;      FILING DATE:  17-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/996,783
;      FILING DATE:  23-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/723,618
;      FILING DATE:  27-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/081,070
;      FILING DATE:  22-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stratford, Carol A.
;      REGISTRATION NUMBER:  34,444
;      REFERENCE/DOCKET NUMBER:  4600-0175.21/G19P3D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  546:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Epstein Barr virus L2  (start site
;      INDIVIDUAL ISOLATE:  90021)
US-08-475-228A-546

  Query Match           88.9%;  Score 16;  DB 2;  Length 50;
  Best Local Similarity  100.0%;  Pred. No. 1.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 5
US-08-482-080A-546/c
; Sequence 546, Application US/08482080A
; Patent No. 6010849
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Cynthia A.
;    APPLICANT:  Cantor, Charles R.
;    APPLICANT:  Andrews, Beth M.
;    APPLICANT:  Turin, Lisa M.
```

```
;    APPLICANT:  Fry, Kirk E.
;    TITLE OF INVENTION:  Sequence-Directed DNA Binding
;    TITLE OF INVENTION:  Molecules, Compositions and Methods
;    NUMBER OF SEQUENCES:  664
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genelabs Technologies, Inc.
;      STREET:  505 Penobscot Drive
;      CITY:  Redwood City
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94063
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,080A
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/171,389
;      FILING DATE:  20-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/123,936
;      FILING DATE:  17-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/996,783
;      FILING DATE:  23-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/723,618
;      FILING DATE:  27-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/081,070
;      FILING DATE:  22-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brady, John F.
;      REGISTRATION NUMBER:  39,118
;      REFERENCE/DOCKET NUMBER:  4600-0175.20/G19P3D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 324-0880
;      TELEFAX:  (650) 324-0960
;  INFORMATION FOR SEQ ID NO:  546:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Epstein Barr virus L2  (start site
;      INDIVIDUAL ISOLATE:  90021)
US-08-482-080A-546

  Query Match             88.9%;  Score 16;  DB 3;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 6
US-09-354-947-546/c
; Sequence 546, Application US/09354947
; Patent No. 6384208
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Cynthia A.
;    APPLICANT:  Cantor, Charles R.
;    APPLICANT:  Andrews, Beth M.
;    APPLICANT:  Turin, Lisa M.
;    APPLICANT:  Fry, Kirk E.
;    TITLE OF INVENTION:  Sequence-Directed DNA Binding
;    TITLE OF INVENTION:  Molecules, Compositions and Methods
;    NUMBER OF SEQUENCES:  664
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genelabs Technologies, Inc.
;      STREET:  505 Penobscot Drive
;      CITY:  Redwood City
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94063
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/354,947
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/482,080
;      FILING DATE:  07-JUN-1995
;      APPLICATION NUMBER:  US 08/171,389
;      FILING DATE:  20-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/123,936
;      FILING DATE:  17-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/996,783
;      FILING DATE:  23-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/723,618
;      FILING DATE:  27-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/081,070
;      FILING DATE:  22-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brady, John F.
;      REGISTRATION NUMBER:  39,118
;      REFERENCE/DOCKET NUMBER:  4600-0175.20/G19P3D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 324-0880
;      TELEFAX:  (650) 324-0960
;  INFORMATION FOR SEQ ID NO:  546:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Epstein Barr virus L2  (start site
;      INDIVIDUAL ISOLATE:  90021)
US-09-354-947-546

  Query Match             88.9%;  Score 16;  DB 4;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 7
PCT-US93-12388-546/c
; Sequence 546, Application PC/TUS9312388
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Sequence-Directed DNA Binding
;    TITLE OF INVENTION:  Molecules, Compositions and Methods
;    NUMBER OF SEQUENCES:  641
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Genelabs Technologies, Inc.
;       STREET:  505 Penobscot Drive
;       CITY:  Redwood City
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94063
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/12388
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/123,936
;       FILING DATE:  17-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/996,783
;       FILING DATE:  23-DEC-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fabian, Gary R.
;       REGISTRATION NUMBER:  33,875
;       REFERENCE/DOCKET NUMBER:  4600-0175.41/G19PCT2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 324-0880
;       TELEFAX:  (415) 324-0960
;   INFORMATION FOR SEQ ID NO:  546:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  50 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  Epstein Barr virus L2 (start site
;       INDIVIDUAL ISOLATE:  90021)
PCT-US93-12388-546

  Query Match             88.9%;  Score 16;  DB 5;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 8
US-08-917-320-18/c
; Sequence 18, Application US/08917320
; Patent No. 5824508
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  No. 5824508 Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/917,320
;      FILING DATE:  25-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1014..3734
US-08-917-320-18

  Query Match             88.9%;  Score 16;  DB 1;  Length 3833;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       3137 GGCTGGTGTCACCTGT 3122


RESULT 9
PCT-US95-04611A-18/c
; Sequence 18, Application PC/TUS9504611A
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  Non Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04611A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1014..3734
PCT-US95-04611A-18

  Query Match             88.9%;  Score 16;  DB 5;  Length 3833;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       3137 GGCTGGTGTCACCTGT 3122


RESULT 10
US-08-783-774-1/c
; Sequence 1, Application US/08783774
; Patent No. 6054130
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Jackman, Winthrop
;    TITLE OF INVENTION:  NON-SPLICING VARIANTS OF
;    TITLE OF INVENTION:  GP350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/783,774
;      FILING DATE:  15-JAN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7682-037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5931 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1014...3734
;      OTHER INFORMATION:
US-08-783-774-1

  Query Match             88.9%;  Score 16;  DB 3;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 0.99;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       3137 GGCTGGTGTCACCTGT 3122


RESULT 11
US-09-556-706B-1/c
; Sequence 1, Application US/09556706B
; Patent No. 6458364
; GENERAL INFORMATION:
;  APPLICANT: Spaete, Richard
;  APPLICANT:  Jackman, Winthrop
;  TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
;  FILE REFERENCE: 7682-050-999
;  CURRENT APPLICATION NUMBER: US/09/556,706B
;  CURRENT FILING DATE:  2000-04-24
;  PRIOR APPLICATION NUMBER: 08/783,774
;  PRIOR FILING DATE: 1997-01-15
;  PRIOR APPLICATION NUMBER: 08/229,291
;  PRIOR FILING DATE: 1994-04-18
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 5931
;   TYPE: DNA
;   ORGANISM: Virus
;   FEATURE:
;   OTHER INFORMATION: gp350/220
US-09-556-706B-1

  Query Match             88.9%;  Score 16;  DB 4;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 0.99;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       3137 GGCTGGTGTCACCTGT 3122


RESULT 12
US-09-171-710-5
; Sequence 5, Application US/09171710
; Patent No. 6323330
; GENERAL INFORMATION:
;  APPLICANT: ISHIDUKA, Yasuyuki
;  APPLICANT: MOCHIZUKI, Reiko
;  TITLE OF INVENTION: NOVEL PROTEINS C16 AND C16N OR GENES ENCODING THE SAME
;  FILE REFERENCE: 0020-4474P
;  CURRENT APPLICATION NUMBER: US/09/171,710
;  CURRENT FILING DATE: 1998-10-23
;  EARLIER APPLICATION NUMBER: PCT/JP97/01391
;  EARLIER FILING DATE: 1997-10-30
;  EARLIER APPLICATION NUMBER: 9-41562
;  EARLIER FILING DATE: 1997-02-10
;  EARLIER APPLICATION NUMBER: 8-127954
;  EARLIER FILING DATE: 1996-04-23
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1808
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-171-710-5

  Query Match             83.3%;  Score 15;  DB 4;  Length 1808;
  Best Local Similarity   100.0%;  Pred. No. 3.8;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db        949 GGCTGGTGTCACCTG 963


RESULT 13
US-09-171-710-3
; Sequence 3, Application US/09171710
; Patent No. 6323330
; GENERAL INFORMATION:
;  APPLICANT: ISHIDUKA, Yasuyuki
```

```
;  APPLICANT: MOCHIZUKI, Reiko
;  TITLE OF INVENTION: NOVEL PROTEINS C16 AND C16N OR GENES ENCODING THE SAME
;  FILE REFERENCE: 0020-4474P
;  CURRENT APPLICATION NUMBER: US/09/171,710
;  CURRENT FILING DATE: 1998-10-23
;  EARLIER APPLICATION NUMBER: PCT/JP97/01391
;  EARLIER FILING DATE: 1997-10-30
;  EARLIER APPLICATION NUMBER: 9-41562
;  EARLIER FILING DATE: 1997-02-10
;  EARLIER APPLICATION NUMBER: 8-127954
;  EARLIER FILING DATE: 1996-04-23
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 3065
;   TYPE: DNA
;   ORGANISM: Mus sp.
US-09-171-710-3

  Query Match             83.3%;  Score 15;  DB 4;  Length 3065;
  Best Local Similarity   100.0%;  Pred. No. 3.7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db       1184 GGCTGGTGTCACCTG 1198


RESULT 14
US-07-745-206A-12
; Sequence 12, Application US/07745206A
; Patent No. 5429921
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Feldman, Daniel
;    TITLE OF INVENTION:  Human Calcium Channel Compositions and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fitch, Even, Tabin & Flannery
;      STREET:  135 S. LaSalle
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/745,206A
;      FILING DATE:  19910815
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feder, Scott B
;      REFERENCE/DOCKET NUMBER:  51504
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-372-7842
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5467 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(144..3164, 3168..3245, 3249..3386, 3390
;      LOCATION:  ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
;      LOCATION:  ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
;      LOCATION:  ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
;      LOCATION:  ..5096, 5100..5306, 5310..5366, 5370..5465)
US-07-745-206A-12

  Query Match             77.8%;  Score 14;  DB 1;  Length 5467;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTGT 16
              ||||||||||||||
Db       4965 CTGGTGTCACCTGT 4978


RESULT 15
US-08-311-363-12
; Sequence 12, Application US/08311363
; Patent No. 5876958
;  GENERAL INFORMATION:
;    APPLICANT:  Harpold, Michael
;    APPLICANT:  Ellis, Steven
;    APPLICANT:  Williams, Mark
;    APPLICANT:  Feldman, Daniel
;    APPLICANT:  McCue, Ann
;    APPLICANT:  Brenner, Robert
;    TITLE OF INVENTION:  Human Calcium Channel Compositions and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Martin, Haller & McClain
;      STREET:  1660 Union Street
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-2926
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/311,363
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/745,206
;      FILING DATE:  15-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seidman, Stephanie L.
;      REGISTRATION NUMBER:  33,779
;      REFERENCE/DOCKET NUMBER:  6362-51506
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619)238-0999
;      TELEFAX:  (619)238-0062
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5467 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(144..3164, 3168..3245, 3249..3386, 3390
;      LOCATION:  ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
;      LOCATION:  ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
;      LOCATION:  ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
;      LOCATION:  ..5096, 5100..5306, 5310..5366, 5370..5465)
US-08-311-363-12

  Query Match             77.8%;  Score 14;  DB 2;  Length 5467;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          3 CTGGTGTCACCTGT 16
              ||||||||||||||
Db       4965 CTGGTGTCACCTGT 4978


RESULT 16
US-09-657-453A-24/c
; Sequence 24, Application US/09657453A
; Patent No. 6458591
; GENERAL INFORMATION:
;  APPLICANT:      Brett P. Monia
;  APPLICANT:  Jacqueline Wyatt
;  TITLE OF INVENTION:      ANTISENSE MODULATION OF PHOSPHORYLASE KINASE ALPHA 2 EXPRESS
;  FILE REFERENCE:      RTS-0136
;  CURRENT APPLICATION NUMBER: US/09/657,453A
;  CURRENT FILING DATE:  2000-09-07
;  NUMBER OF SEQ ID NOS:      105
; SEQ ID NO 24
;    LENGTH: 261
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (148)...(177)
US-09-657-453A-24

  Query Match             72.2%;  Score 13;  DB 4;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 GGTGTCACCTGTT 17
              |||||||||||||
Db        185 GGTGTCACCTGTT 173


RESULT 17
US-09-141-027-9/c
; Sequence 9, Application US/09141027A
; Patent No. 6372454
; GENERAL INFORMATION:
;  APPLICANT: Duan, et al.
;  TITLE OF INVENTION: Follistatin-3
;  FILE REFERENCE: PF388
;  CURRENT APPLICATION NUMBER: US/09/141,027A
;  CURRENT FILING DATE: 1998-08-27
;  EARLIER APPLICATION NUMBER: 60/656,248
;  EARLIER FILING DATE: 1997-08-29
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;    LENGTH: 308
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (3)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (19)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (24)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (29)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (34)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (38)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (40)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (50)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (83)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (107)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (205)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (220)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (237)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (272)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (242)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (297)
;    OTHER INFORMATION: n equals a, t, g, or c
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (308)
;    OTHER INFORMATION: n equals a, t, g, or c
US-09-141-027-9

  Query Match             72.2%;  Score 13;  DB 4;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCT 14
              |||||||||||||
Db         18 GCTGGTGTCACCT 6


RESULT 18
US-09-617-804-9/c
; Sequence 9, Application US/09617804
; Patent No. 6537966
; GENERAL INFORMATION:
;  APPLICANT:  Duan, et al.
;  TITLE OF INVENTION:  Follistatin-3
;  FILE REFERENCE:  PF388P1
;  CURRENT APPLICATION NUMBER: US/09/617,804
;  CURRENT FILING DATE:  2000-07-14
;  PRIOR APPLICATION NUMBER:  60/144,088
;  PRIOR FILING DATE:  1999-07-16
;  PRIOR APPLICATION NUMBER:  09/141,027
;  PRIOR FILING DATE:  1998-08-27
```

```
;  PRIOR APPLICATION NUMBER:  60/056,248
;  PRIOR FILING DATE:  1997-08-29
;  PRIOR APPLICATION NUMBER:  PCT/US98/17710
;  PRIOR FILING DATE:  1998-08-27
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 308
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (19)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (24)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (29)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (34)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (38)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (40)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (50)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (83)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (107)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (205)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (220)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (237)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (272)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (242)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (297)
;   OTHER INFORMATION: n equals a, t, g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (308)
;   OTHER INFORMATION: n equals a, t, g, or c
US-09-617-804-9

  Query Match            72.2%;  Score 13;  DB 4;  Length 308;
  Best Local Similarity  100.0%;  Pred. No. 55;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCT 14
              |||||||||||||
Db         18 GCTGGTGTCACCT 6
```

```
RESULT 19
US-09-702-705-1159/c
; Sequence 1159, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1159
;   LENGTH: 361
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(361)
;   OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1159

  Query Match            72.2%;  Score 13;  DB 4;  Length 361;
  Best Local Similarity  100.0%;  Pred. No. 55;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTG 15
              |||||||||||||
Db        148 CTGGTGTCACCTG 136


RESULT 20
US-09-736-457-1159/c
; Sequence 1159, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1159
;   LENGTH: 361
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(361)
;   OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1159

  Query Match            72.2%;  Score 13;  DB 4;  Length 361;
```

```
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTG 15
              |||||||||||||
Db        148 CTGGTGTCACCTG 136


RESULT 21
US-09-228-986-40
; Sequence 40, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
;  APPLICANT: Strabala, Timothy
;  APPLICANT: Nieuwenhuizen, Niels
;  TITLE OF INVENTION: Compositions Isolated from Plant Cells
;  TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
;  FILE REFERENCE: 11000/1020
;  CURRENT APPLICATION NUMBER: US/09/228,986
;  CURRENT FILING DATE: 1999-01-12
;  NUMBER OF SEQ ID NOS: 130
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 40
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-228-986-40

  Query Match             72.2%;  Score 13;  DB 4;  Length 504;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTG 15
              |||||||||||||
Db        279 CTGGTGTCACCTG 291


RESULT 22
US-09-173-300-10
; Sequence 10, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/09/173,300
;  CURRENT FILING DATE: 1998-10-15
;  EARLIER APPLICATION NUMBER: 60/063,423
;  EARLIER FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 10
;   LENGTH: 794
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-173-300-10

  Query Match             72.2%;  Score 13;  DB 4;  Length 794;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGGTGTCACCTGT 16
              |||||||||||||
Db        437 TGGTGTCACCTGT 449


RESULT 23
US-09-221-456-1/c
; Sequence 1, Application US/09221456
; Patent No. 6162899
; GENERAL INFORMATION:
;   APPLICANT:  SATHE, GANESH
;   APPLICANT:  HALSEY, WENDY
;   APPLICANT:  MUIR, ALISON
;   APPLICANT:  CHAMBERS, JON
;   APPLICANT:  SZEKERES, PHILIP
;   TITLE OF INVENTION:  METHODS OF SCREENING FOR AGONISTS
;   TITLE OF INVENTION:  AND ANTAGONISTS OF THE HNEAA81 RECEPTOR
;   NUMBER OF SEQUENCES:  2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Ratner & Prestia
;     STREET:  P.O. Box 980
;     CITY:  Valley Forge
;     STATE:  PA
;     COUNTRY:  USA
;     ZIP:  19482
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/221,456
;     FILING DATE:  28-DEC-1998
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/956,975
;     FILING DATE:  23-OCT-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Prestia, Paul F
;     REGISTRATION NUMBER:  23,031
;     REFERENCE/DOCKET NUMBER:  GH-70318-1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  610-407-0700
;     TELEFAX:  610-407-0700
;     TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1124 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
US-09-221-456-1

  Query Match             72.2%;  Score 13;  DB 3;  Length 1124;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 GTGTCACCTGTTA 18
              |||||||||||||
Db         89 GTGTCACCTGTTA 77


RESULT 24
US-09-558-740-1/c
; Sequence 1, Application US/09558740
; Patent No. 6358695
; GENERAL INFORMATION:
;  APPLICANT: SATHE, GANESH
;  APPLICANT:  HALSEY, WENDY
;  APPLICANT:  MUIR, ALISON
;  APPLICANT:  CHAMBERS, JON
;  APPLICANT:  SZEKERES, PHILIP
;  TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND
;  TITLE OF INVENTION:  ANTAGONISTS OF THE HNEAA81 RECEPTOR
;  FILE REFERENCE: GH-70318-2
;  CURRENT APPLICATION NUMBER: US/09/558,740
;  CURRENT FILING DATE:  2000-04-26
;  PRIOR APPLICATION NUMBER: 08/956,975
;  PRIOR FILING DATE: 1997-10-23
;  PRIOR APPLICATION NUMBER: 09/221,456
```

```
;  PRIOR FILING DATE: 1998-12-28
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1124
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-558-740-1

  Query Match             72.2%;  Score 13;  DB 4;  Length 1124;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 GTGTCACCTGTTA 18
              |||||||||||||
Db         89 GTGTCACCTGTTA 77


RESULT 25
US-09-328-352-3822/c
; Sequence 3822, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3822
;   LENGTH: 1302
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3822

  Query Match             72.2%;  Score 13;  DB 4;  Length 1302;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGGTGTCACCTGT 16
              |||||||||||||
Db        322 TGGTGTCACCTGT 310


RESULT 26
US-09-328-352-2907/c
; Sequence 2907, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2907
;   LENGTH: 1716
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-2907

  Query Match             72.2%;  Score 13;  DB 4;  Length 1716;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 GTGTCACCTGTTA 18
              |||||||||||||
Db       1328 GTGTCACCTGTTA 1316


RESULT 27
US-09-099-676-2
; Sequence 2, Application US/09099676
; Patent No. 6100075
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
;    TITLE OF INVENTION:  HOMOLOG
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/099,676
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0532 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-855-0572
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1742 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PROSNON01
;      CLONE:  2278458
US-09-099-676-2

  Query Match             72.2%;  Score 13;  DB 3;  Length 1742;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACC 13
              |||||||||||||
Db        479 GGCTGGTGTCACC 491


RESULT 28
US-09-565-910-2
; Sequence 2, Application US/09565910
; Patent No. 6268192
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
;    TITLE OF INVENTION:  HOMOLOG
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
```

```
;       COUNTRY:  USA
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows
;       SOFTWARE:  FastSEQ for Windows Version 2.0b
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/565,910
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  09/099,676
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cerrone, Michael C
;       REGISTRATION NUMBER:  39,132
;       REFERENCE/DOCKET NUMBER:  PF-0532 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-855-0555
;       TELEFAX:  650-855-0572
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1742 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  PROSNON01
;       CLONE:  2278458
US-09-565-910-2

  Query Match             72.2%;  Score 13;  DB 3;  Length 1742;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACC 13
              |||||||||||||
Db        479 GGCTGGTGTCACC 491


RESULT 29
US-08-454-455-3/c
; Sequence 3, Application US/08454455
; Patent No. 5635601
;  GENERAL INFORMATION:
;    APPLICANT:  Moyle, Matthew
;    APPLICANT:  McLean, John W.
;    TITLE OF INVENTION:  NOVEL BETA INTEGRIN SUBUNIT
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      CCUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 720 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,455
;      FILING DATE:  30-May-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/193989
;      FILING DATE:  09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/004142
;      FILING DATE:  13-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/670607
;      FILING DATE:  14-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  P0699C2D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2972 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-454-455-3

  Query Match             72.2%;  Score 13;  DB 1;  Length 2972;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTG 15
              |||||||||||||
Db       1012 CTGGTGTCACCTG 1000


RESULT 30
US-08-454-455-5/c
; Sequence 5, Application US/08454455
; Patent No. 5635601
;  GENERAL INFORMATION:
;    APPLICANT:  Moyle, Matthew
;    APPLICANT:  McLean, John W.
;    TITLE OF INVENTION:  NOVEL BETA INTEGRIN SUBUNIT
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 720 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,455
;      FILING DATE:  30-May-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/193989
;      FILING DATE:  09-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/004142
;      FILING DATE:  13-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/670607
;      FILING DATE:  14-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  P0699C2D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3789 base pairs
```

```
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-454-455-5

  Query Match             72.2%;  Score 13;  DB 1;  Length 3789;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTG 15
              |||||||||||||
Db       1041 CTGGTGTCACCTG 1029


RESULT 31
US-08-474-067-1
; Sequence 1, Application US/08474067
; Patent No. 5811518
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,067
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1682
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3959 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  45..2181
US-08-474-067-1

  Query Match             72.2%;  Score 13;  DB 1;  Length 3959;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGGTGTCACCTGT 16
              |||||||||||||
Db        976 TGGTGTCACCTGT 988


RESULT 32
US-08-474-068A-1
; Sequence 1, Application US/08474068A
; Patent No. 5837525
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,068A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1683
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3959 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  45..2181
US-08-474-068A-1

  Query Match             72.2%;  Score 13;  DB 2;  Length 3959;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGGTGTCACCTGT 16
              |||||||||||||
Db        976 TGGTGTCACCTGT 988


RESULT 33
US-08-472-481-1
; Sequence 1, Application US/08472481
; Patent No. 5863804
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/472,481
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/213,361
;       FILING DATE:  14-MAY-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/607,293
;       FILING DATE:  30-OCT-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LJ 1686
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3959 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  45..2181
US-08-472-481-1

  Query Match             72.2%;  Score 13;  DB 2;  Length 3959;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TGGTGTCACCTGT 16
              |||||||||||||
Db        976 TGGTGTCACCTGT 988


RESULT 34
PCT-US96-01314-39/c
; Sequence 39, Application PC/TUS9601314
;  GENERAL INFORMATION:
;    APPLICANT:  M. Amin Arnaout
;    TITLE OF INVENTION:  METHODS FOR IDENTIFYING INTEGRIN
;    TITLE OF INVENTION:  ANTAGONISTS
;    NUMBER OF SEQUENCES:  78
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/01314
;      FILING DATE:  30-JAN-96
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/380,167
;      FILING DATE:  30-JAN-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John W. Freeman
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00786/267001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US96-01314-39

  Query Match             72.2%;  Score 13;  DB 5;  Length 5137;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACC 13
              |||||||||||||
Db       4208 GGCTGGTGTCACC 4196


RESULT 35
US-08-476-062A-39/c
; Sequence 39, Application US/08476062A
; Patent No. 5877275
;  GENERAL INFORMATION:
;    APPLICANT:  Arnaout, M. Amin
;    TITLE OF INVENTION:  CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
;    TITLE OF INVENTION:  RESPONSES WITH BETA2 INTEGRINS
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,062A
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/216,081
;      FILING DATE:  21-MAR-1994
;      APPLICATION NUMBER:  07/637,830
;      FILING DATE:  04-JAN-1991
;      APPLICATION NUMBER:  07/539,842
;      FILING DATE:  18-JUN-1990
;      APPLICATION NUMBER:  07/212,573
;      FILING DATE:  28-JUN-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00786/068003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5138 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  95...3604
US-08-476-062A-39

  Query Match             72.2%;  Score 13;  DB 2;  Length 5138;
  Best Local Similarity   100.0%;  Pred. No. 50;
```

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCTGGTGTCACC 13
              |||||||||||||
Db       4209 GGCTGGTGTCACC 4197
```

Search completed: August 15, 2003, 11:00:16
Job time : 43.025 secs

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:57:44 ; Search time 113.4 Seconds
(without alignments)
428.482 Million cell updates/sec

Title: US-10-074-620-1
Perfect score: 18
Sequence: 1 ggctggtgtcacctgtta 18

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		18	100.0	18	24	ABA00268	EBNA 2 primer, Pos
2		16	88.9	20	16	AAQ91011	Primer binding to
3	c	16	88.9	50	15	AAQ69796	Epstein Barr virus
4	c	16	88.9	50	18	AAT64258	EBV L2 (start site
5	c	16	88.9	50	20	AAX17546	Test sequence from
6	c	16	88.9	50	24	ABK83037	DNA binding molecu
7	c	16	88.9	435	21	AAC01776	Human secreted pro
8	c	16	88.9	455	24	AAS61522	Lung small cell ca
9		16	88.9	675	22	AAI89104	Human polynucleoti
10	c	16	88.9	1142	19	AAV43611	Human secreted pro
11	c	16	88.9	1152	19	AAV62754	Human secreted pro
12	c	16	88.9	1152	24	ABQ92057	Human polynucleoti
13	c	16	88.9	1217	21	AAC77733	Human cancer assoc
14		16	88.9	1278	20	AAZ17770	Human gene express
15	c	16	88.9	2382	25	ACC46555	Human dithp secret
16	c	16	88.9	2721	6	AAN50114	DNA sequence encod
17	c	16	88.9	5931	16	AAT04821	EBV gp350/220 cDNA
18		16	88.9	13559	22	ABA15144	Human nervous syst
19		15	83.3	65	24	ABN52794	Mouse spliced tran
20		15	83.3	609	23	ABV56544	Human prostate exp
21		15	83.3	616	21	AAF11311	Aspergillus niger
22		15	83.3	1419	22	AAH45451	Murine epilepsy-ca
23		15	83.3	1699	24	ABZ11825	Human polynucleoti
24		15	83.3	1808	18	AAV02313	C16N gene for prom
25		15	83.3	2293	21	AAZ44730	Human C16N-1 cDNA.
26		15	83.3	2293	21	AAZ40179	Human C16N-1 codin
27		15	83.3	2301	21	AAZ44731	Human C16N-2 cDNA.
28		15	83.3	2301	21	AAZ40180	Human C16N-2 codin
29		15	83.3	2952	20	AAZ33584	Human breast tumou
30		15	83.3	3035	21	AAC77831	Human cancer assoc
31		15	83.3	3035	22	AAH34809	Human colon cancer
32		15	83.3	3065	18	AAV02312	C16N gene for prom
33		15	83.3	3337	21	AAZ44729	Murine C16N-2 cDNA
34		15	83.3	3337	21	AAZ40178	Mouse C16N-2 codin
35		15	83.3	3674	21	AAZ44728	Murine C16N-1 cDNA
36		15	83.3	3674	21	AAZ40177	Mouse C16N-1 codin
37		15	83.3	4537	25	ABX63337	Human cDNA #337 di
38		15	83.3	6501	22	AAS06682	Cochliobolus heter
39		15	83.3	6550	24	ABS68409	Fungal peptide syn
40		15	83.3	6553	24	ABS68449	DNA encoding C. he
41		15	83.3	6672	24	ABL68121	Ovary cancer relat
42	c	15	83.3	7702	21	AAA88739	Human protein tyro
43	c	15	83.3	7702	21	AAZ91908	Human protein tyro
44	c	15	83.3	7702	21	AAZ59133	LAR tyrosine phosp
45		15	83.3	7705	22	AAH98405	Human EST-derived
46		15	83.3	7705	22	AAS22684	Human cDNA encodin
47	c	15	83.3	7741	22	AAS22448	Human cDNA encodin
48	c	15	83.3	7945	23	ABV27897	Human prostate exp
49		15	83.3	16831	23	AAS59607	Propionibacterium
50		15	83.3	42115	24	ABS68452	C. heterostrophus
51		14	77.8	249	25	ABZ72979	Rice leaf EST, SEQ
52		14	77.8	327	22	AAI80980	Human polynucleoti
53		14	77.8	370	24	ABL80865	Human ovarian canc
54		14	77.8	383	22	AAH99005	Murine EST-derived
55		14	77.8	411	25	ABX50538	Bovine EST associa
56		14	77.8	454	23	AAS72255	DNA encoding novel
57		14	77.8	455	22	AAH81534	Human differential
58		14	77.8	465	22	ABA52087	Human foetal liver
59		14	77.8	465	22	ABA21899	Probe #365 for gen
60		14	77.8	465	22	AAK00369	Human brain expres
61		14	77.8	465	22	AAK25812	Human bone marrow
62		14	77.8	465	22	AAI10441	Probe #374 for gen
63		14	77.8	465	22	AAI31694	Probe #380 used to
64		14	77.8	465	22	AAI00377	Probe #368 used to
65		14	77.8	465	23	ABS25400	Human liver single
66		14	77.8	465	24	ABS00393	Human genome-deriv
67		14	77.8	1109	24	ABK35284	Human cDNA encodin
68	c	14	77.8	1220	22	AAS46203	Human DNA encoding
69	c	14	77.8	1220	25	ACA57961	Human PRO19626 cDN
70	c	14	77.8	1220	25	ABX98431	Human cDNA encodin
71	c	14	77.8	1220	25	ABX98933	Novel human secret
72	c	14	77.8	1220	25	ACA05978	Human secreted/tra
73	c	14	77.8	1220	25	ABX98022	Human PRO polynucl
74	c	14	77.8	1220	25	ABX78806	Human PRO polynucl
75	c	14	77.8	1220	25	ABX75819	Human cDNA encodin
76	c	14	77.8	1220	25	ABX77024	Human PRO polynucl
77	c	14	77.8	1220	25	ABX16864	Human cDNA encodin
78	c	14	77.8	1548	21	AAC43772	Zea mays DNA fragm
79	c	14	77.8	1855	24	ABS57811	Human cDNA encodin
80		14	77.8	2017	22	AAH99525	Human protein enco
81		14	77.8	2970	25	AAD47350	Human transporter

	82	14	77.8	3069	24	AAD36308 Human transporter
	83	14	77.8	3088	24	ABZ35367 Human gene express
	84	14	77.8	3913	22	AAH17826 Human cDNA sequenc
c	85	14	77.8	4447	23	AAS72881 DNA encoding novel
	86	14	77.8	4614	23	AAS72259 DNA encoding novel
	87	14	77.8	19628	22	AAK71839 Human immune/haema
c	88	14	77.8	32107	22	AAS32249 Human DNA repair a
c	89	14	77.8	32187	24	ABS67552 Novel human DNA re
	90	14	77.8	58837	24	ABK52612 Human Claspin geno
c	91	14	77.8	122888	24	ABK83569 Human cDNA differe
c	92	13	72.2	60	24	ABN40753 Human spliced tran
c	93	13	72.2	145	23	AAS58139 cDNA #815 encoding
c	94	13	72.2	261	24	ABK69167 Human phosphorylas
c	95	13	72.2	294	21	AAC25223 Human secreted pro
c	96	13	72.2	300	20	AAZ14789 Human gene express
c	97	13	72.2	308	20	AAX28130 Human follistatin-
c	98	13	72.2	308	22	AAD02679 HAQAG52R cDNA clon
c	99	13	72.2	350	22	AAK61553 Human immune/haema
c	100	13	72.2	352	22	AAK76174 Human immune/haema
c	101	13	72.2	352	22	AAK76175 Human immune/haema
c	102	13	72.2	361	24	ABK39121 cDNA encoding lung
c	103	13	72.2	361	25	ACA11450 Human lung adenoca
c	104	13	72.2	361	25	ACA02636 Lung cancer therap
	105	13	72.2	369	14	AAQ59915 Human brain Expres
	106	13	72.2	371	24	ABK29789 Colon adenocarcino
	107	13	72.2	375	22	AAS35637 Human cardiovascul
	108	13	72.2	375	22	AAS36753 Human cardiovascul
	109	13	72.2	378	23	AAS58045 cDNA #721 encoding
c	110	13	72.2	413	22	AAS25389 Human ovarian PCR-
	111	13	72.2	414	21	AAC98247 Human colon cancer
c	112	13	72.2	445	23	AAS74105 DNA encoding novel
	113	13	72.2	460	23	AAS84880 DNA encoding novel
	114	13	72.2	495	22	AAS25208 Human ovarian PCR-
	115	13	72.2	495	22	AAH83855 Human ovarian tumo
	116	13	72.2	504	21	AAA79302 Pinus radiata cell
	117	13	72.2	506	22	AAS24619 Human ovarian PCR-
	118	13	72.2	506	22	AAH83242 Human ovarian tumo
	119	13	72.2	527	22	AAL05183 Human reproductive
	120	13	72.2	527	23	ABL98070 Human testicular a

ALIGNMENTS

RESULT 1
ABA00268
ID ABA00268 standard; DNA; 18 BP.
XX
AC ABA00268;
XX
DT 29-NOV-2002 (first entry)
XX
DE EBNA 2 primer, Position 90030:90049.
XX
KW Primer; amplify; PCR; probe; detection; Epstein-Barr virus; EBV; ss.
XX
OS Epstein-barr virus.
XX
PN WO200264842-A2.
XX
PD 22-AUG-2002.
XX
PF 13-FEB-2002; 2002WO-US04339.
XX
PR 13-FEB-2001; 2001US-268439P.
XX
PA (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI Witte DP, Groen PA;
XX
DR WPI: 2002-667015/71.
XX
PT New compositions comprising nucleic acid sequences which specifically
PT hybridizes to Epstein-Barr virus (EBV) nucleic acid, for detecting EBV
PT in clinical specimens to determine patients at high risk of to
PT developing EBV infections -
XX
PS Claim 1; Page 44; 59pp; English.
XX
CC The sequences given in ABA00268-75 are primers and probes which were
CC used in the compositions of the invention for the detection of
CC Epstein-Barr virus (EBV). The compositions comprise at least one
CC purified and isolated oligonucleotide consisting of a nucleic acid
CC sequence which complements and specifically hybridizes to EBV nucleic
CC acid. The oligonucleotide sequences and compositions comprising them
CC are useful for detecting EBV in clinical specimens to determine
CC patients who are at high risk to develop serious and costly medical
CC complications, and allow for better clinical management of these
CC patients by earlier recognition of their infection status. The
CC oligonucleotide sequences may also be used to amplify EBV DNA
CC sequences. The use of the oligonucleotide sequences in the assay for
CC detecting EBV has a broad dynamic range of detection from less than
CC 10-100000000 copies of EBV DNA, is less labour intensive requiring only
CC one reaction tube for the EBV determination, highly sensitive, accurate
CC and has a rapid turn around time with assays that are completed,
CC including amplification, probe specific hybridization, and calculation
CC of copy number in less than 1 hour. The method may be adapted to
CC automated systems.
XX
SQ Sequence 18 BP; 2 A; 4 C; 6 G; 6 T; 0 other;

Query Match 100.0%; Score 18; DB 24; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGTTA 18
||||||||||||||||||
Db 1 GGCTGGTGTCACCTGTTA 18

RESULT 2
AAQ91011
ID AAQ91011 standard; DNA; 20 BP.
XX
AC AAQ91011;
XX
DT 01-FEB-1996 (first entry)
XX
DE Primer binding to 5' end od EBV nuc antigen gene.
XX
KW Primer; PCR; amplification; DNA polymerase; exonuclease; Pfu; Taq;
KW Klenow fragment; T4; T7; Deep Vent; synthesis; mismatch; human; antibody;
KW heavy chain variable region; ss.
XX
OS Synthetic.
XX
PN WO9516028-A1.
XX
PD 15-JUN-1995.
XX
PF 07-DEC-1994; 94WO-US14065.
XX
PR 16-FEB-1994; 94US-0197791.
PR 08-DEC-1993; 93US-0164290.
XX
PA (STRA-) STRATAGENE.
XX
PI Mullinax RL, Sorge JA;
XX
DR WPI; 1995-224316/29.
XX
PT Compsn. useful for polynucleotide synthesis and cyclical
PT amplification - comprising a mixt. contg. an enzyme with 3'-5'
PT exo-nuclease activity and a DNA polymerase with less 3'-5'
PT exo-nuclease activity than the enzyme
XX

PS Examples; Page 35; 66pp; English.
XX
CC Primers AAQ90984-Q91028 are examples of primers for testing a novel
CC composition for polynucleotide synthesis comprising a DNA polymerase
CC with high 3'-5' exonuclease activity in conjunction with a DNA polymerase
CC with less 3'-5' exonuclease activity, pref. Pfu and Taq DNA polymerases
CC respectively. Other DNA polymerases containing high 3'-5' exonuclease
CC activity include E.coli DNA polymerase I, Klenow fragment, T4, T7, Vent
CC or Deep Vent DNA polymerases. The use of a DNA polymerase with high
CC 3'-5' exonuclease activity is designed to overcome the inability of DNA
CC polymerases with low 3'-5' exonuclease activities to initiate synthesis
CC from primers containing 3' terminal mismatches, e.g. due to errors
CC introduced during a PCR cycle.
CC This primer binds to a region at the 5' end of the Epstein-Barr
CC virus nuc antigen gene.
XX
SQ Sequence 20 BP; 2 A; 4 C; 7 G; 7 T; 0 other;

Query Match 88.9%; Score 16; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 1 GGCTGGTGTCACCTGT 16

RESULT 3
AAQ69796/c
ID AAQ69796 standard; DNA; 50 BP.
XX
AC AAQ69796;
XX
DT 25-MAR-2003 (updated)
DT 06-MAR-1995 (first entry)
XX
DE Epstein Barr virus R1 L2 (start site 90021), target region.
XX
KW DNA protein-binding assay; test sequence; screening sequence;
KW promoter; target; TATA box; Herpes Simplex Virus; HSV;
KW origin of replication; UL9; transcription factor; TFIID: ds.
XX
OS Synthetic.
XX
PN WO9414980-A1.
XX
PD 07-JUL-1994.
XX
PF 20-DEC-1993; 93WO-US12388.
XX
PR 23-DEC-1992; 92US-0996783.
PR 17-SEP-1993; 93US-0123936.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
XX
DR WPI; 1994-234711/28.
XX
PT Sequence-directed DNA-binding molecules - useful in
PT pharmaceuticals and as molecular reagents
XX
PS Claim 28; Page 485; 587pp; English.
XX
CC A DNA protein-binding assay is provided, useful for screening
CC libraries of synthetic or biological cpds. for their ability
CC to bind DNA test sequences. The assay is versatile in that any
CC number of test sequences can be tested by placing the test sequence
CC adjacent to a defined protein-binding screening sequence. Binding
CC of mols. to these test sequences changes the binding characteristics
CC of the protein mol. to its cognate binding sequence. When such a mol.
CC binds the test sequence, the equilibrium of the DNA:protein complexes
CC is disturbed, generating changes in the concentration of free DNA probe.
CC One application of this method is to eucaryotic general transcription
CC factors (e.g. TFIID), where the target region is typically selected
CC from DNA sequences adjacent to the binding site for the eucaryotic
CC transcription factor. Numerous exemplary test sequences are given:
CC the sequences in AAQ69251-731 and AAQ69850 correspond to promoter
CC targets (typically, TATA box-contg. sites) for human genes and the
CC sequences in AAQ69732-849 correspond to promoter targets for viral genes.
CC The test sequences may also be randomly generated. DNA:protein
CC interaction may be used for screening purposes, e.g. the Herpes Simplex
CC Virus (HSV) origin of replication and UL9 (see AAQ69851-52, AAQ69865 and
CC AAQ69891).
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 50 BP; 18 A; 14 C; 10 G; 8 T; 0 other;

Query Match 88.9%; Score 16; DB 15; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 41 GGCTGGTGTCACCTGT 26

RESULT 4
AAT64258/c
ID AAT64258 standard; DNA; 50 BP.
XX
AC AAT64258;
XX
DT 25-MAR-2003 (updated)
DT 17-MAR-1997 (first entry)
XX
DE EBV L2 (start site 90021) TFIID binding site.
XX
KW Duplex DNA; target region; binding characteristic; DNA binding protein;
KW TFIID; transcription factor; binding site; inhibition; enhance;
KW cancer; inherited genetic disorder; ds.
XX
OS Epstein-barr virus.
XX
PN US5578444-A.
XX
PD 26-NOV-1996.
XX
PF 20-DEC-1993; 93US-0171389.
XX
PR 20-DEC-1993; 93US-0171389.
PR 27-JUN-1991; 91US-0723618.
PR 23-DEC-1992; 92US-0996783.
PR 17-SEP-1993; 93US-0123936.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
XX
DR WPI; 1997-020402/02.
XX
PT Altering binding characteristics of DNA binding proteins to duplex
PT DNA - by attaching specific small cpd. to target region close to the
PT protein's binding site, useful in treatment of viral disease, cancer
PT etc
XX
PS Claim 6; Column 377; 264pp; English.
XX
CC The sequences given in AAT63713-4312 represent duplex DNA's which act
CC as target regions in the method of the invention. The method for
CC altering the binding characteristics of a DNA-binding protein to duplex
CC DNA comprises contacting the duplex DNA with a small molecule which
CC binds sequence-specifically to a target region, where, when the small
CC molecule is bound to the target region, it is adjacent to, but not
CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
CC The small molecule is added at a concentration effective to alter the

CC binding of the DNA binding protein, pref. TFIID, to its binding site on
CC the duplex DNA. The binding of the small molecule may inhibit or
CC enhance the binding of the DNA-binding protein to its binding site. The
CC compounds isolated using this method are potentially useful as
CC therapeutic agents for treatment of any disease which involves a
CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
CC The method is suitable for screening large biological or chemical
CC libraries and allows determination of sequence-specific and relative
CC affinities of known DNA-binding agents for different DNA sequences.
CC The design of these duplex DNA's allows a single DNA:protein interaction
CC to be used for screening sequence-specific, or preferential, DNA binding
CC proteins that recognise almost any possible sequence (see also AAT49539-
CC 74).
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 50 BP; 18 A; 14 C; 10 G; 8 T; 0 other;

Query Match 88.9%; Score 16; DB 18; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 41 GGCTGGTGTCACCTGT 26

RESULT 5
AAX17546/c
ID AAX17546 standard; DNA; 50 BP.
XX
AC AAX17546;
XX
DT 06-MAY-1999 (first entry)
XX
DE Test sequence from Epstein Barr virus L2 (start site 90021).
XX
KW Test sequence; DNA-binding molecule; screening sequence; human;
KW nucleic acid amplification; target; viral; ds.
XX
OS Epstein-Barr virus.
XX
PN US5869241-A.
XX
PD 09-FEB-1999.
XX
PF 07-JUN-1995; 95US-0475228.
XX
PR 20-DEC-1993; 93US-0171389.
PR 27-JUN-1991; 91US-0723618.
PR 23-DEC-1992; 92US-0996783.
PR 17-SEP-1993; 93US-0123936.
PR 07-JUN-1995; 95US-0475228.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
XX
DR WPI; 1999-152755/13.
XX
PT Determination of DNA sequence preference of a DNA-binding molecule -
PT based on inhibition of binding of protein to oligonucleotide
PT sequence attached to test sequence
XX
PS Claim 3; Columns 379-380; 270pp; English.
XX
CC Sequences AAX17001 to AAX17600 represent specifically claimed target
CC test sequences that are used in the method of the invention of
CC determining the DNA sequence preference of a DNA-binding molecule. The
CC method comprises: (i) adding a test molecule and a DNA-binding protein to
CC a mixture of duplex DNA test oligonucleotides, each of the test
CC oligonucleotides having a test sequence adjacent to a screening sequence,
CC where the screening sequence binds to the DNA-binding protein with a
CC binding affinity that is independent of the DNA sequence of the test
CC sequence, and where the mixture of duplex DNA test oligonucleotides
CC includes several test sequences; (ii) incubating the test molecule, the
CC mixture of duplex DNA test oligonucleotides and the DNA-binding protein
CC for a time sufficient to permit binding of the test molecule to test
CC sequences in the duplex DNA; (iii) separating unbound test
CC oligonucleotides from test oligonucleotides bound to binding protein;
CC (iv) amplifying the unbound test oligonucleotides; (v) repeating steps
CC (ii) to (iv); (vi) isolating the amplified test oligonucleotides; and
CC (vii) sequencing the isolated test oligonucleotides. Test sequences
CC AAX17001-X17481 and AAX17600 correspond to promoter targets for human
CC genes and test sequences AAX17482-X17599 correspond to promoter targets
CC for viral genes.
XX
SQ Sequence 50 BP; 18 A; 14 C; 10 G; 8 T; 0 other;

Query Match 88.9%; Score 16; DB 20; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 41 GGCTGGTGTCACCTGT 26

RESULT 6
ABK83037/c
ID ABK83037 standard; DNA; 50 BP.
XX
AC ABK83037;
XX
DT 27-AUG-2002 (first entry)
XX
DE DNA binding molecule screening method test sequence #546.
XX
KW DNA binding molecule screening; inhibition of transcription;
KW infection; human immunodeficiency virus; HIV; parasite; cancer;
KW cardiovascular; respiratory; gastrointestinal; endocrine; metabolic;
KW rheumatic; immunological; haematological; neurological;
KW psychiatric; dermatological; ophthalmological; musculo-skeletal;
KW urogenital disorder; ss.
XX
OS Synthetic.
XX
PN US6384208-B1.
XX
PD 07-MAY-2002.
XX
PF 15-JUL-1999; 99US-0354947.
XX
PR 20-DEC-1993; 93US-0171389.
PR 07-JUN-1995; 95US-0482080.
PR 27-JUN-1991; 91US-0723618.
PR 23-DEC-1992; 92US-0996783.
PR 17-SEP-1993; 93US-0123936.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Edwards CA, Cantor CR, Andrews BM, Turin LM, Fry KE;
XX
DR WPI; 2002-442819/47.
XX
PT Decreasing transcriptional activity of genes for treating infections or
PT cancer, by administration of an agent that binds to two non-overlapping
PT regions of the gene -
XX
PS Example 15; SEQ ID No 546; 98pp; English.
XX
CC The invention relates to a method of decreasing transcriptional activity
CC in a duplex deoxyribonucleic acid (DNA) template (T1) comprising
CC contacting (T1) with a binding agent comprising at least one small duplex
CC DNA-binding molecule (T2) coupled to at least one other small duplex-
CC binding molecule that binds to a non-overlapping region of target
CC sequence (TS). The method is useful for inhibiting transcription of a

CC range of disease-related genes for treating infections (by viruses,
CC including human immunodeficiency virus, bacteria, fungi, protozoa
CC and parasites), cancer, cardiovascular, respiratory, gastrointestinal,
CC endocrine/metabolic, rheumatic/immunological, haematological,
CC neurological, psychiatric, dermatological, ophthalmological,
CC musculo-skeletal, genetic or urogenital disorders. The method provides
CC sequence-specific inhibition of transcription of pathological genes
CC without affecting transcription of cellular genes regulated by the same
CC transcription factor, and can be applied to regulation of any gene.
CC ABK82492-ABK83155 represent DNA binding molecule test sequences used in
CC the method of the invention.
XX
SQ Sequence 50 BP; 18 A; 14 C; 10 G; 8 T; 0 other;

Query Match 88.9%; Score 16; DB 24; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 41 GGCTGGTGTCACCTGT 26

RESULT 7
AAC01776/c
ID AAC01776 standard; cDNA; 435 BP.
XX
AC AAC01776;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1774.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG01770.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1774; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 435 BP; 105 A; 114 C; 133 G; 79 T; 4 other;

Query Match 88.9%; Score 16; DB 21; Length 435;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 186 GGCTGGTGTCACCTGT 171

RESULT 8
AAS61522/c
ID AAS61522 standard; cDNA; 455 BP.
XX
AC AAS61522;
XX
DT 29-JAN-2002 (first entry)
XX
DE Lung small cell carcinoma antigen, cDNA #63.
XX
KW Human; cytostatic; antitumour; lung small cell cancer antigen;
KW tumour; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200177168-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US11859.
XX
PR 11-APR-2000; 2000US-196780P.
PR 21-JUN-2000; 2000US-213361P.
PR 01-SEP-2000; 2000US-229763P.
PR 05-SEP-2000; 2000US-230629P.
PR 14-SEP-2000; 2000US-232565P.
PR 19-DEC-2000; 2000US-257037P.
PR 08-JAN-2001; 2001US-260796P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
DR WPI; 2002-010896/01.
XX
PT Lung tumour polynucleotide and polypeptides useful in therapy and
PT diagnosis of cancer especially lung cancer -
XX
PS Claim 1; Page 146; 295pp; English.
XX
CC The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilising oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridised to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient. (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumour protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for
CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer in a patient. An
CC isolated T cell population is useful for removing tumour cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. AAS61460-AAS61874 represent novel human lung small cell
CC cancer antigen coding sequences of the invention.
XX
SQ Sequence 455 BP; 122 A; 116 C; 138 G; 78 T; 1 other;

```
  Query Match             88.9%;  Score 16;  DB 24;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        177 GGCTGGTGTCACCTGT 162


RESULT 9
AAI89104
ID   AAI89104 standard; cDNA; 675 BP.
XX
AC   AAI89104;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 9164.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO09173.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 9164; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 675 BP; 189 A; 165 C; 147 G; 172 T; 2 other;

  Query Match             88.9%;  Score 16;  DB 22;  Length 675;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTGTTA 18
              ||||||||||||||||
Db        437 CTGGTGTCACCTGTTA 452


RESULT 10
AAV43611/c
ID   AAV43611 standard; DNA; 1142 BP.
XX
AC   AAV43611;
XX
DT   24-SEP-1998  (first entry)
XX
DE   Human secreted protein 11 encoding DNA.
XX
KW   Secreted protein; human; cell proliferation; cytokine activity;
KW   tissue growth; cellular differentiation; regeneration; activin;
KW   inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW   anti-inflammatory activity; biomarker; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..788
FT                   /*tag=  a
FT                   /product= "human secreted protein"
XX
PN   WO9825959-A2.
XX
PD   18-JUN-1998.
XX
PF   11-DEC-1997;   97WO-US22787.
XX
PR   11-DEC-1996;   96US-0032757.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Escobedo J,  Garcia P,  Hu Q,  Kothakota S,  Williams LT;
XX
DR   WPI; 1998-348453/30.
DR   P-PSDB; AAW63691.
XX
PT   Secreted human polypeptides - having cytokine, cell proliferation or
PT   differentiation, activin or inhibin, tumour inhibition or
PT   anti-inflammatory activities
XX
PS   Claim 6; Page 38; 78pp; English.
XX
CC   This DNA encodes a human secreted protein. The specification provides
CC   secreted protein sequences (AAW63681 to AAW63699) encoded by the nucleic
CC   acid sequences shown in AAV43601 to AAV43619. The invention provides a
CC   method of identifying a secreted polypeptide which is modified by rough
CC   microsomes. The secreted proteins can be used in assays to determine
CC   biological activities, such as cytokine, cell proliferation, or cellular
CC   differentiation activities, tissue growth or regeneration, activin or
CC   inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC   thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC   anti-inflammatory activity.  The proteins can also be used as
CC   biomarkers, to identify tissues or cell types which express the proteins,
CC   or a stage- or disease-specific alteration in protein expression. They
CC   can be used in protein interaction assays, to identify ligands or binding
CC   proteins. Compounds which affect the biological activities of the
CC   secreted proteins or their ability to interact with specific ligands can
CC   be identified using the proteins in screening assays. The proteins and
CC   antibodies that bind specifically to the protein can also be used to
CC   design diagnostic tests and therapeutic compositions for diseases which
CC   may be associated with altered expression of these proteins. Fusion
CC   proteins comprising, e.g. signal sequences or transmembrane domains of
CC   the proteins can be used to target other protein domains to cellular
CC   membrane or they can be secreted extracellularly.
XX
SQ   Sequence 1142 BP; 304 A; 335 C; 331 G; 172 T; 0 other;

  Query Match             88.9%;  Score 16;  DB 19;  Length 1142;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        159 GGCTGGTGTCACCTGT 144
```

RESULT 11
AAV62754/c
ID AAV62754 standard; cDNA; 1152 BP.
XX
AC AAV62754;
XX
DT 15-FEB-1999 (first entry)
XX
DE Human secreted protein clone fm150_1 cDNA.
XX
KW Secreted protein; human; fm150_1; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   11..973
FT   /*tag= a
XX
PN WO9846757-A2.
XX
PD 22-OCT-1998.
XX
PF 14-APR-1998; 98WO-US07999.
XX
PR 13-APR-1998; 98US-0059487.
PR 15-APR-1997; 97US-0843374.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-568731/48.
DR P-PSDB; AAW74728.
XX
PT New polynucleotide(s) encoding secreted human proteins - are derived
PT from, e.g. human foetal brain or foetal kidney cDNA libraries,
PT potentially useful as, e.g. vaccines or thrombolytic agents
XX
PS Claim 34; Page 95; 120pp; English.
XX
CC Full-length cDNA clone fm150_1 includes an open reading frame
CC encoding a human secreted protein (see AAW74728). It was isolated
CC from a human adult brain cDNA library using methods which are
CC selective for cDNAs encoding secreted proteins, or was identified
CC as encoding a secreted or transmembrane protein on the basis of
CC computer analysis of the amino acid sequence of the encoded protein.
CC It shows some similarity to database sequences. The invention
CC provides polynucleotides (see AAV62746-55) from human foetal brain,
CC adult testis, adult brain, adult kidney and foetal kidney (all
CC deposited as composite clone ATCC 98404), which encode human
CC secreted proteins (see AAW74720-29). The polynucleotides and
CC proteins are predicted to have biological activities which would
CC make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data are
CC given. Suggested activities include nutritional activity, immune
CC stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity,
CC haemostatic and thrombolytic activity, receptor/ligand activity,
CC antiinflammatory activity, cadherin/tumour invasion suppressor
CC activity, and tumour inhibition activity. The polynucleotides are
CC also stated to be useful for gene therapy, and can be used in
CC recombinant production of the polypeptidess.
XX
SQ Sequence 1152 BP; 342 A; 329 C; 317 G; 164 T; 0 other;

Query Match 88.9%; Score 16; DB 19; Length 1152;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        137 GGCTGGTGTCACCTGT 122
```

RESULT 12
ABQ92057/c
ID ABQ92057 standard; cDNA; 1152 BP.
XX
AC ABQ92057;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 54.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002065394-A1.
XX
PD 30-MAY-2002.
XX
PF 22-DEC-2000; 2000US-0745763.
XX
PR 18-MAR-1998; 98US-0040963.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2002-582343/62.
DR P-PSDB; ABP61843.
XX
PT Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX
PS Claim 189; Page 195-196; 284pp; English.
XX
CC The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC activities. (I) can be used to manipulate stem cells in culture to give
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and

CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polynucleotide of the invention.
XX
SQ Sequence 1152 BP; 342 A; 329 C; 317 G; 164 T; 0 other;

Query Match 88.9%; Score 16; DB 24; Length 1152;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 137 GGCTGGTGTCACCTGT 122

RESULT 13
AAC77733/c
ID AAC77733 standard; cDNA; 1217 BP.
XX
AC AAC77733;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:127.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR P-PSDB; AAB43524.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 709; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 1217 BP; 300 A; 360 C; 367 G; 187 T; 3 other;

Query Match 88.9%; Score 16; DB 21; Length 1217;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 235 GGCTGGTGTCACCTGT 220

RESULT 14
AAZ17770
ID AAZ17770 standard; cDNA; 1278 BP.
XX
AC AAZ17770;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:5243.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX

PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2475-2476; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 1278 BP; 216 A; 383 C; 387 G; 282 T; 10 other;

Query Match 88.9%; Score 16; DB 20; Length 1278;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 1118 GGCTGGTGTCACCTGT 1133

RESULT 15
ACC46555/c
ID ACC46555 standard; cDNA; 2382 BP.
XX
AC ACC46555;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human dithp secreted/extracellular matrix protein-encoding cDNA.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW secreted protein; extracellular matrix; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200297031-A2.
XX
PD 05-DEC-2002.
XX
PF 27-MAR-2002; 2002WO-US10056.
XX
PR 28-MAR-2001; 2001US-279619P.
PR 29-MAR-2001; 2001US-280067P.
PR 29-MAR-2001; 2001US-280068P.
PR 16-MAY-2001; 2001US-291280P.
PR 17-MAY-2001; 2001US-291829P.
PR 17-MAY-2001; 2001US-291849P.
PR 19-JUN-2001; 2001US-299428P.
PR 20-JUN-2001; 2001US-299776P.
PR 20-JUN-2001; 2001US-300001P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
DR WPI; 2003-129518/12.
DR P-PSDB; ABR41618.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for
PT identifying test compound which specifically binds to a polypeptide
PT encoded by human diagnostic and therapeutic polynucleotide, and to
PT induce antibodies -
XX
PS Claim 2; SEQ ID No 476; 591pp; English.
XX
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their
CC encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates
CC to polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods
CC of detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a dithp cDNA encoding a DITHP protein
CC which is a secreted/extracellular matrix protein.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2382 BP; 456 A; 1032 C; 337 G; 556 T; 1 other;

Query Match 88.9%; Score 16; DB 25; Length 2382;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 1869 GGCTGGTGTCACCTGT 1854

RESULT 16
AAN50114/c
ID AAN50114 standard; DNA; 2721 BP.
XX
AC AAN50114;
XX
DT 25-MAR-2003 (updated)
DT 17-OCT-1991 (first entry)
XX
DE DNA sequence encoding Epstein-Barr virus (EBV) outer surface protein.
XX

```
KW   Epstein-Barr virus; antigen; vaccine; ss.
XX
OS   Epstein-Barr virus.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1..2721
FT                   /*tag=  a
FT                   /label= EBV surface protein antigen
XX
PN   EP151079-A.
XX
PD   07-AUG-1985.
XX
PF   28-JAN-1985;   85EP-0400141.
XX
PR   23-JUL-1984;   84US-0633558.
PR   30-JAN-1984;   84US-0575352.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Kleff E,  Tanner J,  Hummel M,  Belsel C;
XX
DR   WPI; 1985-191978/32.
DR   P-PSDB; AAP50073.
XX
PT   New fragment of Epstein-Barr Virus DNA - useful in vector to
PT   express polypeptide for use in prepn. of vaccine against the
PT   virus and for use in diagnosis.
XX
PS   Claim 1; Page 21-23; 26pp; English.
XX
CC   The sequence encodes an outer surface viral protein of EBV, used
CC   to generate antibodies reacting with the surface proteins of
CC   EBV-infected cells, and in the preparation of a vaccine against EBV.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2721 BP; 762 A; 876 C; 557 G; 526 T; 0 other;

  Query Match             88.9%;  Score 16;  DB 6;  Length 2721;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       2124 GGCTGGTGTCACCTGT 2109


RESULT 17
AAT04821/c
ID   AAT04821 standard; cDNA; 5931 BP.
XX
AC   AAT04821;
XX
DT   18-JAN-1996  (first entry)
XX
DE   EBV gp350/220 cDNA.
XX
KW   EBV; gp350; gp220; gp350/gp220; non-splicing variant; vaccine; ds.
XX
OS   Epstein-Barr virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1014..3737
FT                   /*tag=  a
FT   sig_peptide     1014..1067
FT                   /*tag=  b
FT   mat_peptide     1068..3734
FT                   /*tag=  c
FT   misc_feature    2514..2515
FT                   /*tag=  d
FT                   /function= splice donor site
FT                   /note= "bases 2513-2517 (AAGT) are replaced by
FT                          GTCA in the non-splicing variant"
FT   misc_feature    3105..3106
FT                   /*tag=  e
FT                   /function= splice acceptor site
FT                   /note= "bases 3104-3107 (AGGT) are replaced by
FT                          TGGA in the non-splicing variant"
FT   polyA_signal    3742..3747
FT                   /*tag=  f
XX
PN   WO9528488-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;   95WO-US04611.
XX
PR   18-APR-1994;   94US-0229291.
XX
PA   (AVIR-) AVIRON.
XX
PI   Jackman WT,  Spaete R;
XX
DR   WPI; 1995-373802/48.
DR   P-PSDB; AAR80144.
XX
PT   New DNA encoding a homogeneous gp350 protein - can be used for
PT   preventing and treating Epstein-Barr virus-related diseases or
PT   conditions
XX
PS   Claim 2; Fig.1; 61pp; English.
XX
CC   The donor and acceptor splice sites of the EBV gene encoding gp350/
CC   220 are mutated by replacement of native nucleotides by non-native
CC   nucleotides, without altering the encoded amino acid sequence,
CC   resulting in elimination of gp220 prodn.  Recombinant homogeneous
CC   gp350, useful in vaccines, is expressed in mammalian or insect cell
CC   hosts.
XX
SQ   Sequence 5931 BP; 1453 A; 1782 C; 1437 G; 1259 T; 0 other;

  Query Match             88.9%;  Score 16;  DB 16;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       3137 GGCTGGTGTCACCTGT 3122


RESULT 18
ABA15144
ID   ABA15144 standard; DNA; 13559 BP.
XX
AC   ABA15144;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7475.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
```

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -

XX
PS Disclosure; SEQ ID NO 7475; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 13559 BP; 2882 A; 3867 C; 4239 G; 2571 T; 0 other;

Query Match 88.9%; Score 16; DB 22; Length 13559;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 12761 GGCTGGTGTCACCTGT 12776

RESULT 19
ABN52794
ID ABN52794 standard; DNA; 65 BP.
XX
AC ABN52794;
XX
DT 15-JUL-2002 (first entry)
XX
DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:25542.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Mus musculus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 25542; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 65 BP; 8 A; 11 C; 23 G; 23 T; 0 other;

Query Match 83.3%; Score 15; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
|||||||||||||||
Db 8 GGCTGGTGTCACCTG 22

RESULT 20
ABV56544
ID ABV56544 standard; cDNA; 609 BP.
XX
AC ABV56544;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 56535.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 10906; 11750pp; English.
XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 609 BP; 186 A; 116 C; 99 G; 208 T; 0 other;

Query Match 83.3%; Score 15; DB 23; Length 609;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CTGGTGTCACCTGTT 17
||||||||||||||||
Db 45 CTGGTGTCACCTGTT 59

RESULT 21
AAF11311
ID AAF11311 standard; cDNA; 616 BP.
XX
AC AAF11311;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus niger EST SEQ ID NO:3834.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus niger.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 87; Page 1716; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 616 BP; 138 A; 179 C; 141 G; 146 T; 12 other;

Query Match 83.3%; Score 15; DB 21; Length 616;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
|||||||||||||||
Db 244 GGCTGGTGTCACCTG 258

RESULT 22
AAH45451
ID AAH45451 standard; cDNA; 1419 BP.
XX
AC AAH45451;
XX
DT 06-SEP-2001 (first entry)
XX
DE Murine epilepsy-causing gene Epm2A cDNA sequence.
XX
KW Mouse; Epm2A; epilepsy; transgenic animal; knockout mouse; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 67..1059
FT /*tag= a
FT /product= "Epm2A"
FT /note= "Protein tyrosine phosphatase"
XX
PN JP2001095587-A.
XX
PD 10-APR-2001.
XX
PF 01-OCT-1999; 99JP-0281632.
XX
PR 01-OCT-1999; 99JP-0281632.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
DR WPI; 2001-341250/36.
DR P-PSDB; AAG62454.
XX
PT Mouse epilepsy-causing gene for the analysis of epilepsy and use in a
PT mouse model of epilepsy -
XX
PS Claim 2; Page 6-8; 10pp; Japanese.
XX
CC This invention relates to a murine gene (Epm2A) and its encoded protein,
CC which cause epilepsy. Epm2A is a protein tyrosine phosphatase. Included
CC in the invention is a method of preparing an Epm2A knockout mouse. The
CC gene can be used for the analysis of Epm2A expression and to create an
CC epilepsy model animal. The present sequence represents cDNA encoding
CC murine epilepsy causing Epm2A protein.

```
XX
SQ   Sequence 1419 BP; 293 A; 376 C; 437 G; 313 T; 0 other;

  Query Match            83.3%;  Score 15;  DB 22;  Length 1419;
  Best Local Similarity  100.0%;  Pred. No. 21;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db       1146 GGCTGGTGTCACCTG 1160


RESULT 23
ABZ11825
ID   ABZ11825 standard; cDNA; 1699 BP.
XX
AC   ABZ11825;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 707.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69608.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
PS   Claim 1; SEQ ID NO 707; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1699 BP; 512 A; 296 C; 287 G; 604 T; 0 other;

  Query Match            83.3%;  Score 15;  DB 24;  Length 1699;
  Best Local Similarity  100.0%;  Pred. No. 21;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTGTT 17
              |||||||||||||||
Db        742 CTGGTGTCACCTGTT 756


RESULT 24
AAV02313
ID   AAV02313 standard; cDNA; 1808 BP.
XX
AC   AAV02313;
XX
DT   06-MAY-1998  (first entry)
XX
DE   C16N gene for promoting neuron survival and type 1 collagen production.
XX
KW   C16; C16N; neuron survival; type 1 collagen; calcium regulation;
KW   hypercalcaemia; hypertension; diabetes; arteriosclerosis; cancer;
KW   myocardial infarction; hydroxyapatite; osteoblast; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1740
FT                   /*tag=  a
FT                   /product= "C16N"
XX
PN   WO9740150-A1.
XX
PD   30-OCT-1997.
XX
PF   23-APR-1997;   97WO-JP01391.
XX
PR   10-FEB-1997;   97JP-0041562.
PR   23-APR-1996;   96JP-0127954.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI   Ishiduka Y,  Mochizuki R;
XX
DR   WPI; 1997-535834/49.
DR   P-PSDB; AAW31366.
XX
PT   Proteins C16 and C16N promote neuron survival and type 1 collagen
PT   production - for treatment of diseases involving collagen
PT   production, calcium regulation or neuron survival
XX
PS   Claim 6; Page 61-62; 86pp; Japanese.
XX
CC   The present sequence encodes C16N which can: (a) induce differentiation
CC   into cells which can degrade hydroxyapatite; (b) maintain neuron
CC   survival; (c) inhibit osteoblast proliferation; and (d) promote type 1
CC   collagen expression in osteoblasts. C16 and C16N are agents for the
CC   treatment of a broad range of diseases including hypercalcaemia,
CC   hypertension, diabetes, arteriosclerosis, myocardial infarction and
CC   terminal cancer. They may also be used as a screen for potential
CC   inhibitors of their activity for possible medicinal use. Transgenic
CC   animals containing DNA coding for the proteins can be used as model
CC   organisms and for the production of recombinant C16/C16N.
XX
SQ   Sequence 1808 BP; 336 A; 548 C; 551 G; 373 T; 0 other;

  Query Match            83.3%;  Score 15;  DB 18;  Length 1808;
  Best Local Similarity  100.0%;  Pred. No. 21;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db        949 GGCTGGTGTCACCTG 963


RESULT 25
AAZ44730
ID   AAZ44730 standard; cDNA; 2293 BP.
XX
AC   AAZ44730;
XX
DT   17-APR-2000  (first entry)
XX
DE   Human C16N-1 cDNA.
XX
KW   C16N; C16N-1; C16N-2; human; cartilage cell differentiation; therapy;
KW   cartilage disorder treatment agent; arthritis; rheumatoid arthritis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..2293
FT                   /*tag=  a
FT                   /product= "C16N-1"
FT                   /note= "no stop codon given"
XX
PN   WO200001405-A1.
XX
PD   13-JAN-2000.
XX
PF   02-JUL-1999;   99WO-JP03577.
XX
PR   06-JUL-1998;   98JP-0190889.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI   Ishiduka Y,  Mochizuki R;
XX
DR   WPI; 2000-126903/11.
DR   P-PSDB; AAY51326.
XX
PT   Cartilage cell differentiation promoters or remedies for cartilage
PT   failure  -
XX
PS   Example 2; Page 36-41; 55pp; Japanese.
XX
CC   This invention describes a novel cartilage cell differentiation
CC   promoter, or cartilage disorder treatment agent, which contains C16N,
CC   C16N-1, C16N-2, or their protein analogs, or one of the genes encoding
CC   them, as active ingredient. The promoters and cartilage disorder
CC   treatment agents are for the therapy of cartilage failure including
CC   deformans arthritis, diseases due to cartilage formation abnormality,
CC   deletion of cartilage in bone fraction, joint cartilage caused by injury,
CC   or damage to articular disc, acute purulent arthritis, tuberculous
CC   arthritis, syphilous arthritis, chronic rheumatoid arthritis, rheumatic
CC   fever, systemic erythematosus, deformative spinal diseases, and
CC   intervertebral hernia. This sequence encodes the human C16N-1 protein
CC   which is described in the method of the invention.
XX
SQ   Sequence 2293 BP; 417 A; 716 C; 691 G; 469 T; 0 other;

  Query Match             83.3%;  Score 15;  DB 21;  Length 2293;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db       1505 GGCTGGTGTCACCTG 1519


RESULT 26
AAZ40179
ID   AAZ40179 standard; cDNA; 2293 BP.
XX
AC   AAZ40179;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human C16N-1 coding sequence.
XX
KW   C16N-1; C16N-2; human; myeloid cell differentiation; neuron;
KW   hydroxyapatite decomposition; osteoblast growth inhibitor;
KW   type I collagen expression; ds.
XX
OS   Homo sapiens.
XX
PN   JP11308995-A.
XX
PD   09-NOV-1999.
XX
PF   28-APR-1998;   98JP-0134440.
XX
PR   28-APR-1998;   98JP-0134440.
XX
PA   (SUMU ) SUMITOMO SEIYAKU KK.
XX
DR   WPI; 2000-046933/04.
DR   P-PSDB; AAY55017.
XX
PT   New proteins C16N-1 and C16N-2 or a gene coding them - can have
PT   activity limited to specific tissue or site
XX
PS   Claim 2; Page 18-21; 30pp; Japanese.
XX
CC   This sequence encodes a C16N-1 protein of the invention. The invention
CC   also relates to C16N-2 proteins. The proteins of the invention have the
CC   following features: (1) it has a differentiation inducing activity,
CC   inducing differentiation of a myeloid cell to a cell having
CC   hydroxyapatite decomposing activity; (2) it has an activity of
CC   maintaining the survival of neuron; (3) it has an activity of inhibiting
CC   growth of osteoblasts; and (4) it has an activity of promoting expression
CC   of type I collagen in osteoblasts. The activity of the proteins can be
CC   limited to a specific tissue or a specific site.
XX
SQ   Sequence 2293 BP; 417 A; 716 C; 691 G; 469 T; 0 other;

  Query Match             83.3%;  Score 15;  DB 21;  Length 2293;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              |||||||||||||||
Db       1505 GGCTGGTGTCACCTG 1519


RESULT 27
AAZ44731
ID   AAZ44731 standard; cDNA; 2301 BP.
XX
AC   AAZ44731;
XX
DT   17-APR-2000  (first entry)
XX
DE   Human C16N-2 cDNA.
XX
KW   C16N; C16N-1; C16N-2; human; cartilage cell differentiation; therapy;
KW   cartilage disorder treatment agent; arthritis; rheumatoid arthritis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..2233
FT                   /*tag=  a
FT                   /product= "C16N-2"
XX
PN   WO200001405-A1.
```

XX
PD 13-JAN-2000.
XX
PF 02-JUL-1999; 99WO-JP03577.
XX
PR 06-JUL-1998; 98JP-0190889.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Ishiduka Y, Mochizuki R;
XX
DR WPI; 2000-126903/11.
DR P-PSDB; AAY51327.
XX
PT Cartilage cell differentiation promoters or remedies for cartilage
PT failure -
XX
PS Example 2; Page 45-50; 55pp; Japanese.
XX
CC This invention describes a novel cartilage cell differentiation
CC promoter, or cartilage disorder treatment agent, which contains C16N,
CC C16N-1, C16N-2, or their protein analogs, or one of the genes encoding
CC them, as active ingredient. The promoters and cartilage disorder
CC treatment agents are for the therapy of cartilage failure including
CC deformans arthritis, diseases due to cartilage formation abnormality,
CC deletion of cartilage in bone fraction, joint cartilage caused by injury,
CC or damage to articular disc, acute purulent arthritis, tuberculous
CC arthritis, syphilous arthritis, chronic rheumatoid arthritis, rheumatic
CC fever, systemic erythematosus, deformative spinal diseases, and
CC intervertebral hernia. This sequence encodes the human C16N-2 protein
CC which is described in the method of the invention.
XX
SQ Sequence 2301 BP; 425 A; 714 C; 703 G; 459 T; 0 other;

Query Match 83.3%; Score 15; DB 21; Length 2301;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
||||||||||||||
Db 1442 GGCTGGTGTCACCTG 1456

RESULT 28
AAZ40180
ID AAZ40180 standard; cDNA; 2301 BP.
XX
AC AAZ40180;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human C16N-2 coding sequence.
XX
KW C16N-1; C16N-2; human; myeloid cell differentiation; neuron;
KW hydroxyapatite decomposition; osteoblast growth inhibitor;
KW type I collagen expression; ds.
XX
OS Homo sapiens.
XX
PN JP11308995-A.
XX
PD 09-NOV-1999.
XX
PF 28-APR-1998; 98JP-0134440.
XX
PR 28-APR-1998; 98JP-0134440.
XX
PA (SUMU) SUMITOMO SEIYAKU KK.
XX
DR WPI; 2000-046933/04.
DR P-PSDB; AAY55018.
XX
PT New proteins C16N-1 and C16N-2 or a gene coding them - can have
PT activity limited to specific tissue or site
XX
PS Claim 2; Page 23-26; 30pp; Japanese.
XX
CC This sequence encodes a C16N-2 protein of the invention. The invention
CC also relates to C16N-1 proteins. The proteins of the invention have the
CC following features: (1) it has a differentiation inducing activity,
CC inducing differentiation of a myeloid cell to a cell having
CC hydroxyapatite decomposing activity; (2) it has an activity of
CC maintaining the survival of neuron; (3) it has an activity of inhibiting
CC growth of osteoblasts; and (4) it has an activity of promoting expression
CC of type I collagen in osteoblasts. The activity of the proteins can be
CC limited to a specific tissue or a specific site.
XX
SQ Sequence 2301 BP; 425 A; 714 C; 703 G; 459 T; 0 other;

Query Match 83.3%; Score 15; DB 21; Length 2301;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
||||||||||||||
Db 1442 GGCTGGTGTCACCTG 1456

RESULT 29
AAZ33584
ID AAZ33584 standard; cDNA; 2952 BP.
XX
AC AAZ33584;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human breast tumour-associated EST 44.
XX
KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS Homo sapiens.
XX
PN DE19813835-A1.
XX
PD 23-SEP-1999.
XX
PF 20-MAR-1998; 98DE-1013835.
XX
PR 20-MAR-1998; 98DE-1013835.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
DR WPI; 1999-528979/45.
XX
PT Human nucleic acid sequences and protein products from normal breast
PT tissue, useful for breast cancer therapy -
XX
PS Claim 3; 132-133; 206pp; German.
XX
CC This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytostatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated
CC with fat metabolism. AAZ33541-Z33610 represent expressed sequence tags
CC described in the method of the invention.
XX
SQ Sequence 2952 BP; 925 A; 564 C; 582 G; 881 T; 0 other;

Query Match 83.3%; Score 15; DB 20; Length 2952;
Best Local Similarity 100.0%; Pred. No. 20;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 CTGGTGTCACCTGTT 17
              ||||||| ||||||||
Db       2836 CTGGTGTCACCTGTT 2850
```

RESULT 30
AAC77831
ID AAC77831 standard; cDNA; 3035 BP.
XX
AC AAC77831;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:225.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR P-PSDB; AAB43622.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 796-797; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 3035 BP; 955 A; 575 C; 588 G; 912 T; 5 other;

Query Match 83.3%; Score 15; DB 21; Length 3035;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 CTGGTGTCACCTGTT 17
              ||||||||||||||
Db       2771 CTGGTGTCACCTGTT 2785
```

RESULT 31
AAH34809
ID AAH34809 standard; cDNA; 3035 BP.
XX
AC AAH34809;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1891.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 2; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG75404.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3405-3406; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 3035 BP; 955 A; 575 C; 590 G; 912 T; 3 other;

Query Match 83.3%; Score 15; DB 22; Length 3035;
Best Local Similarity 100.0%; Pred. No. 20;.
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 CTGGTGTCACCTGTT 17
```

```
              ||||||||||||||||
Db       2771 CTGGTGTCACCTGTT 2785


RESULT 32
AAV02312
ID   AAV02312 standard; cDNA; 3065 BP.
XX
AC   AAV02312;
XX
DT   06-MAY-1998  (first entry)
XX
DE   C16N gene for promoting neuron survival and type 1 collagen production.
XX
KW   C16; C16N; neuron survival; type 1 collagen; calcium regulation;
KW   hypercalcaemia; hypertension; diabetes; arteriosclerosis; cancer;
KW   myocardial infarction; hydroxyapatite; osteoblast; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             236..1975
FT                   /*tag=  a
FT                   /product= "C16N"
XX
PN   WO9740150-A1.
XX
PD   30-OCT-1997.
XX
PF   23-APR-1997;   97WO-JP01391.
XX
PR   10-FEB-1997;   97JP-0041562.
PR   23-APR-1996;   96JP-0127954.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI   Ishiduka Y,  Mochizuki R;
XX
DR   WPI; 1997-535834/49.
DR   P-PSDB; AAW31364.
XX
PT   Proteins C16 and C16N promote neuron survival and type 1 collagen
PT   production - for treatment of diseases involving collagen
PT   production, calcium regulation or neuron survival
XX
PS   Claim 6; Page 55-57; 86pp; Japanese.
XX
CC   The present sequence encodes C16N which can: (a) induce differentiation
CC   into cells which can degrade hydroxyapatite; (b) maintain neuron
CC   survival; (c) inhibit osteoblast proliferation; and (d) promote type 1
CC   collagen expression in osteoblasts. C16 and C16N are agents for the
CC   treatment of a broad range of diseases including hypercalcaemia,
CC   hypertension, diabetes, arteriosclerosis, myocardial infarction and
CC   terminal cancer. They may also be used as a screen for potential
CC   inhibitors of their activity for possible medicinal use. Transgenic
CC   animals containing DNA coding for the proteins can be used as model
CC   organisms and for the production of recombinant C16/C16N.
XX
SQ   Sequence 3065 BP; 571 A; 950 C; 889 G; 655 T; 0 other;

  Query Match             83.3%;  Score 15;  DB 18;  Length 3065;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              ||||||||||||||||
Db       1184 GGCTGGTGTCACCTG 1198


RESULT 33
AAZ44729
ID   AAZ44729 standard; cDNA; 3337 BP.
XX
AC   AAZ44729;
XX
DT   17-APR-2000  (first entry)
XX
DE   Murine C16N-2 cDNA.
XX
KW   C16N; C16N-1; C16N-2; murine; cartilage cell differentiation; therapy;
KW   cartilage disorder treatment agent; arthritis; rheumatoid arthritis; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             109..2247
FT                   /*tag=  a
FT                   /product= "C16N-2"
XX
PN   WO200001405-A1.
XX
PD   13-JAN-2000.
XX
PF   02-JUL-1999;   99WO-JP03577.
XX
PR   06-JUL-1998;   98JP-0190889.
XX
PA   (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI   Ishiduka Y,  Mochizuki R;
XX
DR   WPI; 2000-126903/11.
DR   P-PSDB; AAY51325.
XX
PT   Cartilage cell differentiation promoters or remedies for cartilage
PT   failure -
XX
PS   Example 1; Page 27-33; 55pp; Japanese.
XX
CC   This invention describes a novel cartilage cell differentiation
CC   promoter, or cartilage disorder treatment agent, which contains C16N,
CC   C16N-1, C16N-2, or their protein analogs, or one of the genes encoding
CC   them, as active ingredient. The promoters and cartilage disorder
CC   treatment agents are for the therapy of cartilage failure including
CC   deformans arthritis, diseases due to cartilage formation abnormality,
CC   deletion of cartilage in bone fraction, joint cartilage caused by injury,
CC   or damage to articular disc, acute purulent arthritis, tuberculous
CC   arthritis, syphilous arthritis, chronic rheumatoid arthritis, rheumatic
CC   fever, systemic erythematosus, deformative spinal diseases, and
CC   intervertebral hernia. This sequence encodes the murine C16N-2 protein
CC   which is described in the method of the invention.
XX
SQ   Sequence 3337 BP; 633 A; 1037 C; 976 G; 691 T; 0 other;

  Query Match             83.3%;  Score 15;  DB 21;  Length 3337;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTG 15
              ||||||||||||||||
Db       1456 GGCTGGTGTCACCTG 1470


RESULT 34
AAZ40178
ID   AAZ40178 standard; cDNA; 3337 BP.
XX
AC   AAZ40178;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Mouse C16N-2 coding sequence.
XX
KW   C16N-1; C16N-2; mouse; myeloid cell differentiation; neuron;
KW   hydroxyapatite decomposition; osteoblast growth inhibitor;
KW   type I collagen expression; ds.
```

XX
OS Mus sp.
XX
PN JP11308995-A.
XX
PD 09-NOV-1999.
XX
PF 28-APR-1998; 98JP-0134440.
XX
PR 28-APR-1998; 98JP-0134440.
XX
PA (SUMU) SUMITOMO SEIYAKU KK.
XX
DR WPI; 2000-046933/04.
DR P-PSDB; AAY55016.
XX
PT New proteins C16N-1 and C16N-2 or a gene coding them - can have
PT activity limited to specific tissue or site
XX
PS Claim 2; Page 13-16; 30pp; Japanese.
XX
CC This sequence encodes a C16N-2 protein of the invention. The invention
CC also relates to C16N-1 proteins. The proteins of the invention have the
CC following features: (1) it has a differentiation inducing activity,
CC inducing differentiation of a myeloid cell to a cell having
CC hydroxyapatite decomposing activity; (2) it has an activity of
CC maintaining the survival of neuron; (3) it has an activity of inhibiting
CC growth of osteoblasts; and (4) it has an activity of promoting expression
CC of type I collagen in osteoblasts. The activity of the proteins can be
CC limited to a specific tissue or a specific site.
XX
SQ Sequence 3337 BP; 633 A; 1037 C; 976 G; 691 T; 0 other;

Query Match 83.3%; Score 15; DB 21; Length 3337;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
||||||||||||||||
Db 1456 GGCTGGTGTCACCTG 1470

RESULT 35
AAZ44728
ID AAZ44728 standard; cDNA; 3674 BP.
XX
AC AAZ44728;
XX
DT 17-APR-2000 (first entry)
XX
DE Murine C16N-1 cDNA.
XX
KW C16N; C16N-1; C16N-2; murine; cartilage cell differentiation; therapy;
KW cartilage disorder treatment agent; arthritis; rheumatoid arthritis; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 395..2584
FT /*tag= a
FT /product= "C16N-1"
XX
PN WO200001405-A1.
XX
PD 13-JAN-2000.
XX
PF 02-JUL-1999; 99WO-JP03577.
XX
PR 06-JUL-1998; 98JP-0190889.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Ishiduka Y, Mochizuki R;
XX
DR WPI; 2000-126903/11.
DR P-PSDB; AAY51324.
XX
PT Cartilage cell differentiation promoters or remedies for cartilage
PT failure -
XX
PS Example 2; Page 18-24; 55pp; Japanese.
XX
CC This invention describes a novel cartilage cell differentiation
CC promoter, or cartilage disorder treatment agent, which contains C16N,
CC C16N-1, C16N-2, or their protein analogs, or one of the genes encoding
CC them, as active ingredient. The promoters and cartilage disorder
CC treatment agents are for the therapy of cartilage failure including
CC deformans arthritis, diseases due to cartilage formation abnormality,
CC deletion of cartilage in bone fraction, joint cartilage caused by injury,
CC or damage to articular disc, acute purulent arthritis, tuberculous
CC arthritis, syphilous arthritis, chronic rheumatoid arthritis, rheumatic
CC fever, systemic erythematosus, deformative spinal diseases, and
CC intervertebral hernia. This sequence encodes the murine C16N-1 protein
CC which is described in the method of the invention.
XX
SQ Sequence 3674 BP; 695 A; 1146 C; 1044 G; 789 T; 0 other;

Query Match 83.3%; Score 15; DB 21; Length 3674;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTG 15
||||||||||||||||
Db 1793 GGCTGGTGTCACCTG 1807

Search completed: August 14, 2003, 21:41:17
Job time : 116.4 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:44:31 ; Search time 32.25 Seconds
(without alignments)
273.726 Million cell updates/sec

Title: US-10-074-620-2
Perfect score: 20
Sequence: 1 ccttaggaggaacaagtccc 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20	100.0	20	1	US-08-197-791-29	Sequence 29, Appl
2	20	100.0	3833	1	US-08-917-320-18	Sequence 18, Appl
3	20	100.0	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
4	20	100.0	5931	3	US-08-783-774-1	Sequence 1, Appli
5	20	100.0	5931	4	US-09-556-706B-1	Sequence 1, Appli
6	16	80.0	6572	4	US-09-620-312D-823	Sequence 823, App
7	14	70.0	854	4	US-09-439-313-354	Sequence 354, App
8	14	70.0	854	4	US-09-352-616A-354	Sequence 354, App
c 9	14	70.0	1245	1	US-08-599-171A-17	Sequence 17, Appl
c 10	14	70.0	1245	2	US-08-646-590B-17	Sequence 17, Appl
c 11	14	70.0	1245	3	US-09-069-226-17	Sequence 17, Appl
c 12	14	70.0	1245	3	US-09-412-184-17	Sequence 17, Appl
c 13	14	70.0	1585	4	US-09-232-160-6	Sequence 6, Appli
c 14	14	70.0	4249	1	US-08-480-784-21	Sequence 21, Appl
c 15	14	70.0	4249	1	US-08-483-553-21	Sequence 21, Appl
c 16	14	70.0	4249	1	US-08-487-002-21	Sequence 21, Appl
c 17	14	70.0	4249	1	US-08-483-554B-21	Sequence 21, Appl
c 18	14	70.0	4249	1	US-08-488-011B-21	Sequence 21, Appl
c 19	14	70.0	4249	3	US-08-850-727-21	Sequence 21, Appl
c 20	14	70.0	4249	5	PCT-US95-10202-21	Sequence 21, Appl
c 21	14	70.0	4249	5	PCT-US95-10203-21	Sequence 21, Appl
c 22	14	70.0	4249	5	PCT-US95-10220-21	Sequence 21, Appl
c 23	13	65.0	126	3	US-08-746-411A-9	Sequence 9, Appli
c 24	13	65.0	126	4	US-08-857-046A-9	Sequence 9, Appli
c 25	13	65.0	126	4	US-09-573-252-9	Sequence 9, Appli
c 26	13	65.0	243	1	US-08-248-474-23	Sequence 23, Appl
c 27	13	65.0	243	3	US-08-756-849-23	Sequence 23, Appl
c 28	13	65.0	450	4	US-09-702-705-1692	Sequence 1692, Ap
c 29	13	65.0	450	4	US-09-736-457-1692	Sequence 1692, Ap
c 30	13	65.0	1525	2	US-08-244-205-4	Sequence 4, Appli
c 31	13	65.0	1525	5	PCT-US92-10284-4	Sequence 4, Appli
c 32	13	65.0	1645	5	PCT-US94-01321-9	Sequence 9, Appli
c 33	13	65.0	1970	4	US-09-389-956-79	Sequence 79, Appl
c 34	13	65.0	2517	4	US-10-020-079-39	Sequence 39, Appl
c 35	13	65.0	2556	4	US-10-020-079-37	Sequence 37, Appl
c 36	13	65.0	2592	4	US-10-020-079-31	Sequence 31, Appl
c 37	13	65.0	2631	4	US-10-020-079-29	Sequence 29, Appl
c 38	13	65.0	2856	4	US-10-020-079-35	Sequence 35, Appl
c 39	13	65.0	2874	4	US-10-020-079-33	Sequence 33, Appl
c 40	13	65.0	2931	4	US-10-020-079-27	Sequence 27, Appl
c 41	13	65.0	2949	4	US-10-020-079-25	Sequence 25, Appl
c 42	13	65.0	3645	2	US-08-663-112-1	Sequence 1, Appli
43	13	65.0	16995	4	US-08-961-527-82	Sequence 82, Appl
44	13	65.0	32207	2	US-08-770-379-20	Sequence 20, Appl
45	13	65.0	32207	3	US-08-757-669A-20	Sequence 20, Appl
46	13	65.0	32207	4	US-09-230-371A-20	Sequence 20, Appl
47	13	65.0	176373	3	US-09-128-155-17	Sequence 17, Appl
c 48	13	65.0	176373	3	US-09-128-155-17	Sequence 17, Appl
c 49	13	65.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c 50	12	60.0	22	2	US-08-506-864A-10	Sequence 10, Appl
c 51	12	60.0	22	2	US-08-851-968-10	Sequence 10, Appl
52	12	60.0	226	3	US-08-906-791-6	Sequence 6, Appli
53	12	60.0	276	4	US-09-499-203-50	Sequence 50, Appl
c 54	12	60.0	281	2	US-08-506-864A-3	Sequence 3, Appli
c 55	12	60.0	281	2	US-08-851-968-3	Sequence 3, Appli
c 56	12	60.0	282	4	US-09-205-258-120	Sequence 120, App
57	12	60.0	284	3	US-08-906-791-7	Sequence 7, Appli
c 58	12	60.0	289	4	US-09-313-294A-6286	Sequence 6286, Ap
59	12	60.0	302	4	US-09-313-294A-1134	Sequence 1134, Ap
60	12	60.0	303	4	US-09-313-294A-4838	Sequence 4838, Ap
c 61	12	60.0	323	2	US-08-506-864A-4	Sequence 4, Appli
c 62	12	60.0	323	2	US-08-851-968-4	Sequence 4, Appli
c 63	12	60.0	354	4	US-09-702-705-1618	Sequence 1618, Ap
c 64	12	60.0	354	4	US-09-736-457-1618	Sequence 1618, Ap
c 65	12	60.0	420	3	US-09-532-803-5	Sequence 5, Appli
c 66	12	60.0	421	3	US-09-211-631-12	Sequence 12, Appl
c 67	12	60.0	421	3	US-09-265-628-12	Sequence 12, Appl
c 68	12	60.0	421	3	US-09-001-141-10	Sequence 10, Appl
c 69	12	60.0	421	4	US-09-653-403-13	Sequence 13, Appl
70	12	60.0	421	4	US-09-643-597-326	Sequence 326, App
c 71	12	60.0	421	4	US-10-013-784-13	Sequence 13, Appl
72	12	60.0	421	4	US-09-480-884A-326	Sequence 326, App
73	12	60.0	421	4	US-09-542-615A-326	Sequence 326, App
74	12	60.0	421	4	US-09-606-421B-326	Sequence 326, App
c 75	12	60.0	474	2	US-08-619-542B-45	Sequence 45, Appl
76	12	60.0	497	4	US-09-589-287B-9	Sequence 9, Appli
77	12	60.0	497	4	US-09-588-947A-9	Sequence 9, Appli
78	12	60.0	504	4	US-09-107-532A-3133	Sequence 3133, Ap
c 79	12	60.0	516	1	US-08-510-878-2	Sequence 2, Appli
c 80	12	60.0	612	1	US-08-756-299-1	Sequence 1, Appli
c 81	12	60.0	612	2	US-08-964-494-1	Sequence 1, Appli
82	12	60.0	627	3	US-09-385-982-201	Sequence 201, App
83	12	60.0	642	4	US-09-107-532A-1927	Sequence 1927, Ap
c 84	12	60.0	669	4	US-09-465-901-19	Sequence 19, Appl
c 85	12	60.0	669	4	US-09-465-901-23	Sequence 23, Appl
c 86	12	60.0	669	4	US-09-465-901-27	Sequence 27, Appl
87	12	60.0	735	4	US-09-107-532A-1185	Sequence 1185, Ap
c 88	12	60.0	748	1	US-08-510-878-3	Sequence 3, Appli
89	12	60.0	768	1	US-08-592-126-116	Sequence 116, App
90	12	60.0	768	4	US-09-168-595-116	Sequence 116, App
c 91	12	60.0	800	3	US-08-998-416-390	Sequence 390, App
92	12	60.0	819	4	US-10-083-304-3	Sequence 3, Appli
93	12	60.0	823	4	US-08-858-207A-235	Sequence 235, App
94	12	60.0	840	3	US-09-264-419C-3	Sequence 3, Appli
c 95	12	60.0	846	2	US-08-619-542B-43	Sequence 43, Appl
c 96	12	60.0	914	4	US-09-177-650-123	Sequence 123, App
c 97	12	60.0	941	4	US-09-312-283C-351	Sequence 351, App
98	12	60.0	999	1	US-08-469-649-1	Sequence 1, Appli
99	12	60.0	999	4	US-09-347-878-59	Sequence 59, Appl
100	12	60.0	1016	2	US-08-930-617-1	Sequence 1, Appli

101	12	60.0	1037	4	US-09-535-008-57	Sequence 57, Appl
102	12	60.0	1049	4	US-09-280-116-176	Sequence 176, App
103	12	60.0	1056	1	US-08-402-217A-1	Sequence 1, Appli
104	12	60.0	1056	1	US-08-700-178-1	Sequence 1, Appli
105	12	60.0	1056	3	US-08-995-654-1	Sequence 1, Appli
106	12	60.0	1059	4	US-09-107-532A-1282	Sequence 1282, Ap
107	12	60.0	1164	4	US-09-107-532A-2348	Sequence 2348, Ap
c 108	12	60.0	1164	4	US-09-107-532A-2348	Sequence 2348, Ap
109	12	60.0	1176	4	US-09-107-532A-555	Sequence 555, App
110	12	60.0	1189	2	US-08-450-042A-2	Sequence 2, Appli
111	12	60.0	1190	4	US-09-390-207-1	Sequence 1, Appli
112	12	60.0	1207	3	US-09-264-419C-1	Sequence 1, Appli
113	12	60.0	1248	4	US-09-134-001C-1980	Sequence 1980, Ap
c 114	12	60.0	1260	4	US-09-206-166-8	Sequence 8, Appli
115	12	60.0	1325	2	US-08-632-470-50	Sequence 50, Appl
116	12	60.0	1339	1	US-07-936-163-2	Sequence 2, Appli
117	12	60.0	1353	1	US-08-756-299-3	Sequence 3, Appli
118	12	60.0	1353	2	US-08-964-494-3	Sequence 3, Appli
119	12	60.0	1380	1	US-07-936-163-1	Sequence 1, Appli
120	12	60.0	1392	6	5463025-2	Patent No. 5463025

ALIGNMENTS

```
RESULT 1
US-08-197-791-29
; Sequence 29, Application US/08197791
; Patent No. 5556772
;  GENERAL INFORMATION:
;    APPLICANT:  Sorge, Joseph A.
;    APPLICANT:  Mullinax, Rebecca L.
;    TITLE OF INVENTION:  NOVEL POLYMERASE COMPOSITIONS AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Limbach and Limbach
;      STREET:  2001 Ferry Building
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/197,791
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/164,290
;      FILING DATE:  08-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bortner, Scott R.
;      REGISTRATION NUMBER:  34,298
;      REFERENCE/DOCKET NUMBER:  STRG 20270 USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-433-4150
;      TELEFAX:  415-433-8716
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-197-791-29

  Query Match             100.0%;  Score 20;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db          1 CCTTAGGAGGAACAAGTCCC 20


RESULT 2
US-08-917-320-18
; Sequence 18, Application US/08917320
; Patent No. 5824508
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  No. 5824508 Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/917,320
;      FILING DATE:  25-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1014..3734
US-08-917-320-18

  Query Match             100.0%;  Score 20;  DB 1;  Length 3833;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 3
PCT-US95-04611A-18
; Sequence 18, Application PC/TUS9504611A
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  Non Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
```

```
;       STREET:  5 Palo Alto Square
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94306
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patent In Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/04611A
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/229,291
;       FILING DATE:  April 18, 1994
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Luann Cserr
;       REGISTRATION NUMBER:  31,822
;       REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-843-5163
;       TELEFAX:  415-857-0663
;       TELEX:  380816 CooleyPA
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3833 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1014..3734
PCT-US95-04611A-18

  Query Match             100.0%;  Score 20;  DB 5;  Length 3833;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 4
US-08-783-774-1
; Sequence 1, Application US/08783774
; Patent No. 6054130
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard
;    APPLICANT:  Jackman, Winthrop
;    TITLE OF INVENTION:  NON-SPLICING VARIANTS OF
;    TITLE OF INVENTION:  GP350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/783,774
;      FILING DATE:  15-JAN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7682-037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5931 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1014...3734
;      OTHER INFORMATION:
US-08-783-774-1

  Query Match             100.0%;  Score 20;  DB 3;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 5
US-09-556-706B-1
; Sequence 1, Application US/09556706B
; Patent No. 6458364
; GENERAL INFORMATION:
;  APPLICANT: Spaete, Richard
;  APPLICANT:  Jackman, Winthrop
;  TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
;  FILE REFERENCE: 7682-050-999
;  CURRENT APPLICATION NUMBER: US/09/556,706B
;  CURRENT FILING DATE:  2000-04-24
;  PRIOR APPLICATION NUMBER: 08/783,774
;  PRIOR FILING DATE: 1997-01-15
;  PRIOR APPLICATION NUMBER: 08/229,291
;  PRIOR FILING DATE: 1994-04-18
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 5931
;   TYPE: DNA
;   ORGANISM: Virus
;   FEATURE:
;   OTHER INFORMATION: gp350/220
US-09-556-706B-1

  Query Match             100.0%;  Score 20;  DB 4;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 0.0087;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 6
US-09-620-312D-823
; Sequence 823, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
```

```
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  John Tillinghast
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 823
;   LENGTH: 6572
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (521)..(2611)
US-09-620-312D-823

  Query Match             80.0%;  Score 16;  DB 4;  Length 6572;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db       4524 CTTAGGAGGAACAAGT 4539


RESULT 7
US-09-439-313-354
; Sequence 354, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 354
;   LENGTH: 854
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-354

  Query Match             70.0%;  Score 14;  DB 4;  Length 854;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTC 18
              ||||||||||||||
Db        497 AGGAGGAACAAGTC 510


RESULT 8
US-09-352-616A-354
; Sequence 354, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Mitcham, Jennifer Lynn
;  TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF PROSTATE CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE:  210121.427C8
;  CURRENT APPLICATION NUMBER: US/09/352,616A
;  CURRENT FILING DATE:  1999-07-13
;  NUMBER OF SEQ ID NOS:  472
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 354
;   LENGTH: 854
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-352-616A-354

  Query Match             70.0%;  Score 14;  DB 4;  Length 854;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTC 18
              ||||||||||||||
Db        497 AGGAGGAACAAGTC 510


RESULT 9
US-08-599-171A-17/c
; Sequence 17, Application US/08599171A
; Patent No. 5814473
;  GENERAL INFORMATION:
;    APPLICANT:  WARREN, Patrick V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,171A
;      FILING DATE:  Concurrently
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HERRON, CHARLES J.
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-38
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1245 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  GENOMIC DNA
US-08-599-171A-17

  Query Match            70.0%;  Score 14;  DB 1;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 19;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAG 16
              ||||||||||||||
Db       1069 TTAGGAGGAACAAG 1056


RESULT 10
US-08-646-590B-17/c
; Sequence 17, Application US/08646590B
; Patent No. 5962283
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Patrick V.
;    APPLICANT:  Swanson, Ronald V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,590B
;      FILING DATE:  08-May-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/599,171
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/01094
;      FILING DATE:  21-January-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/0170O1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1245 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...1242
US-08-646-590B-17

  Query Match            70.0%;  Score 14;  DB 2;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 19;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAG 16
              ||||||||||||||
Db       1069 TTAGGAGGAACAAG 1056


RESULT 11
US-09-069-226-17/c
; Sequence 17, Application US/09069226
; Patent No. 6013509
;  GENERAL INFORMATION:
;    APPLICANT:  WARREN, Patrick V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/069,226
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/599,171
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HERRON, CHARLES J.
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-38
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1245 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  GENOMIC DNA
US-09-069-226-17

  Query Match            70.0%;  Score 14;  DB 3;  Length 1245;
  Best Local Similarity  100.0%;  Pred. No. 19;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAG 16
              ||||||||||||||
Db       1069 TTAGGAGGAACAAG 1056


RESULT 12
US-09-412-184-17/c
; Sequence 17, Application US/09412184
; Patent No. 6268188
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Patrick V.
;    APPLICANT:  Swanson, Ronald V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
```

```
;       STATE:  CA
;       COUNTRY:  US
;       ZIP:  92037
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/412,184
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/646,590
;       FILING DATE:  08-May-1996
;       APPLICATION NUMBER:  08/599,171
;       FILING DATE:  09-FEB-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US97/01094
;       FILING DATE:  21-January-1997
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Haile, Ph.D., Lisa A.
;       REGISTRATION NUMBER:  38,347
;       REFERENCE/DOCKET NUMBER:  09010/0170O1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619/678-5070
;       TELEFAX:  619/678-5099
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1245 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  1...1242
US-09-412-184-17

  Query Match             70.0%;  Score 14;  DB 3;  Length 1245;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAG 16
              ||||||||||||||
Db       1069 TTAGGAGGAACAAG 1056


RESULT 13
US-09-232-160-6/c
; Sequence 6, Application US/09232160
; Patent No. 6368794
; GENERAL INFORMATION:
;  APPLICANT: Steve Daniel
;  APPLICANT: James Gilmore
;  APPLICANT: Susan G. Stuart
;  APPLICANT: Laura Stuve
;  TITLE OF INVENTION: DETECTION OF ALTERED EXPRESSION OF GENES REGULATING CELL
;  TITLE OF INVENTION: PROLIFERATION
;  FILE REFERENCE: PA-0003 US
;  CURRENT APPLICATION NUMBER: US/09/232,160
;  CURRENT FILING DATE: 1999-01-15
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PERL Program
; SEQ ID NO 6
;   LENGTH: 1585
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 1698542
US-09-232-160-6

  Query Match             70.0%;  Score 14;  DB 4;  Length 1585;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GAGGAACAAGTCCC 20
              ||||||||||||||
Db        536 GAGGAACAAGTCCC 523


RESULT 14
US-08-480-784-21/c
; Sequence 21, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
;   APPLICANT:  Skolnick, Mark H.
;   APPLICANT:  Goldgar, David E.
;   APPLICANT:  Miki, Yoshio
;   APPLICANT:  Swenson, Jeff
;   APPLICANT:  Kamb, Alexander
;   APPLICANT:  Harshman, Keith D.
;   APPLICANT:  Shattuck-Eidens, Donna M.
;   APPLICANT:  Tavtigian, Sean V.
;   APPLICANT:  Wiseman, Roger W.
;   APPLICANT:  Futreal, P. Andrew
;   TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;   TITLE OF INVENTION:  Susceptibility Gene
;   NUMBER OF SEQUENCES:  85
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;     STREET:  1201 New York Avenue, N.W., Suite 1000
;     CITY:  Washington
;     STATE:  DC
;     COUNTRY:  USA
;     ZIP:  20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/480,784
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/409,305
;     FILING DATE:  24-MAR-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/348,824
;     FILING DATE:  29-NOV-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/308,104
;     FILING DATE:  16-SEP-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/300,266
;     FILING DATE:  02-SEP-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/289,221
;     FILING DATE:  12-AUG-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Ihnen, Jeffrey L.
;     REGISTRATION NUMBER:  28,957
;     REFERENCE/DOCKET NUMBER:  24884-109347
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  202-962-4810
;     TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  4249 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-480-784-21

  Query Match             70.0%;  Score 14;  DB 1;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 15
US-08-483-553-21/c
; Sequence 21, Application US/08483553
; Patent No. 5709999
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,553
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-483-553-21

  Query Match             70.0%;  Score 14;  DB 1;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 16
US-08-487-002-21/c
; Sequence 21, Application US/08487002
; Patent No. 5710001
;  GENERAL INFORMATION:
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Emi, Mitsuru
;    APPLICANT:  Nakamura, Yusuke
;    APPLICANT:  Durocher, Francine
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,002
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-487-002-21

  Query Match             70.0%;  Score 14;  DB 1;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 17
US-08-483-554B-21/c
; Sequence 21, Application US/08483554B
; Patent No. 5747282
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,554B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-483-554B-21

  Query Match             70.0%;  Score 14;  DB 1;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 18
US-08-488-011B-21/c
; Sequence 21, Application US/08488011B
; Patent No. 5753441
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,011B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-488-011B-21

  Query Match             70.0%;  Score 14;  DB 1;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 19
US-08-850-727-21/c
; Sequence 21, Application US/08850727
; Patent No. 6162897
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,727
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/483,554
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-850-727-21

  Query Match             70.0%;  Score 14;  DB 3;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 20
PCT-US95-10202-21/c
; Sequence 21, Application PC/TUS9510202
;  GENERAL INFORMATION:
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Simard, Jacques
;    APPLICANT:  Emi, Mitsuru
;    APPLICANT:  Nakamura, Yusuke
;    APPLICANT:  Durocher, Francine
;    TITLE OF INVENTION:  In Vivo Mutations and Polymorphisms
;    TITLE OF INVENTION:  in the 17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10202
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08-308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
PCT-US95-10202-21

  Query Match             70.0%;  Score 14;  DB 5;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 21
PCT-US95-10203-21/c
; Sequence 21, Application PC/TUS9510203
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  17q-Linked Breast and Ovarian Cancer
;    TITLE OF INVENTION:  Susceptibility Gene
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10203
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,305
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,824
;      FILING DATE:  29-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08-308,104
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/300,266
;      FILING DATE:  02-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/289,221
;      FILING DATE:  12-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ihnen, Jeffrey L.
;      REGISTRATION NUMBER:  28,957
;      REFERENCE/DOCKET NUMBER:  24884-109347
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-962-4810
;      TELEFAX:  202-962-8300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
PCT-US95-10203-21

  Query Match             70.0%;  Score 14;  DB 5;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 22
PCT-US95-10220-21/c
; Sequence 21, Application PC/TUS9510220
;  GENERAL INFORMATION:
;    APPLICANT:  Skolnick, Mark H.
;    APPLICANT:  Goldgar, David E.
;    APPLICANT:  Miki, Yoshio
;    APPLICANT:  Swenson, Jeff
;    APPLICANT:  Kamb, Alexander
;    APPLICANT:  Harshman, Keith D.
;    APPLICANT:  Shattuck-Eidens, Donna M.
;    APPLICANT:  Tavtigian, Sean V.
;    APPLICANT:  Wiseman, Roger W.
;    APPLICANT:  Futreal, P. Andrew
;    TITLE OF INVENTION:  Method for Diagnosing a
;    TITLE OF INVENTION:  Predisposition for Breast and Ovarian Cancer
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Venable, Baetjer, Howard & Civiletti, LLP
;      STREET:  1201 New York Avenue, N.W., Suite 1000
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10220
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US
;       FILING DATE:  07-JUN-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/409,305
;       FILING DATE:  24-MAR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/348,824
;       FILING DATE:  29-NOV-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08-308,104
;       FILING DATE:  16-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/300,266
;       FILING DATE:  02-SEP-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/289,221
;       FILING DATE:  12-AUG-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ihnen, Jeffrey L.
;       REGISTRATION NUMBER:  28,957
;       REFERENCE/DOCKET NUMBER:  24884-109347
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-962-4810
;       TELEFAX:  202-962-8300
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4249 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
PCT-US95-10220-21

  Query Match             70.0%;  Score 14;  DB 5;  Length 4249;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db       3770 CCTTAGGAGGAACA 3757


RESULT 23
US-08-746-411A-9/c
; Sequence 9, Application US/08746411A
; Patent No. 6117632
;  GENERAL INFORMATION:
;    APPLICANT:  O'Mahony, Daniel J
;    TITLE OF INVENTION:  Peptides Which Enhance Transport Across
;    TITLE OF INVENTION:  Tissues and Methods of Identifying and Using the Same
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary L. Severson, Ph.D., Esq.
;      STREET:  1300 Gould Drive
;      CITY:  Gainesville
;      STATE:  GA
;      COUNTRY:  USA
;      ZIP:  30504
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/746,411A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/006461
;      FILING DATE:  10-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  IE 950865
;      FILING DATE:  10-NOV-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Severson, Mary L
;      REGISTRATION NUMBER:  34,927
;      REFERENCE/DOCKET NUMBER:  96.1060.US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  770 534-8239
;      TELEFAX:  770 534-8247
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "individual isolate"
US-08-746-411A-9

  Query Match             65.0%;  Score 13;  DB 3;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAG 16
              |||||||||||||
Db         37 TAGGAGGAACAAG 25


RESULT 24
US-08-857-046A-9/c
; Sequence 9, Application US/08857046A
; Patent No. 6361938
;  GENERAL INFORMATION:
;    APPLICANT:  O'Mahony, Daniel J
;    APPLICANT:  Alverez, Vernon L
;    APPLICANT:  Seveso, Michela
;    TITLE OF INVENTION:  Peptides Which Enhance Transport Across
;    TITLE OF INVENTION:  Tissues and Methods of Identifying and Using the Same
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary L. Severson, Ph.D., Esq.
;      STREET:  1300 Gould Drive
;      CITY:  Gainesville
;      STATE:  GA
;      COUNTRY:  USA
;      ZIP:  30504
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/857,046A
;      FILING DATE:  15-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/006461
;      FILING DATE:  10-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  IE 950864
;      FILING DATE:  10-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/746,411
;      FILING DATE:  08-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/IE96/00073
```

```
;      FILING DATE:  11-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/IE96/00072
;      FILING DATE:  11-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Severson, Mary L
;      REGISTRATION NUMBER:  34,927
;      REFERENCE/DOCKET NUMBER:  97.1061.US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  770 534-8239
;      TELEFAX:  770 534-8247
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "individual isolate"
US-08-857-046A-9

  Query Match             65.0%;  Score 13;  DB 4;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAG 16
              |||||||||||||
Db         37 TAGGAGGAACAAG 25


RESULT 25
US-09-573-252-9/c
; Sequence 9, Application US/09573252
; Patent No. 6521737
;  GENERAL INFORMATION:
;       APPLICANT: O'Mahony, Daniel J
;       TITLE OF INVENTION: Peptides Which Enhance Transport Across
;                           Tissues and Methods of Identifying and Using the Same
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Mary L. Severson, Ph.D., Esq.
;            STREET: 1300 Gould Drive
;            CITY: Gainesville
;            STATE: GA
;            COUNTRY: USA
;            ZIP: 30504
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/573,252
;            FILING DATE: 19-Aug-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/746,411
;            FILING DATE: 08-NOV-1996
;            APPLICATION NUMBER: US 60/006461
;            FILING DATE: 10-NOV-1995
;            APPLICATION NUMBER: IE 950865
;            FILING DATE: 10-NOV-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Severson, Mary L
;            REGISTRATION NUMBER: 34,927
;            REFERENCE/DOCKET NUMBER: 96.1060.US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 770 534-8239
;            TELEFAX: 770 534-8247
;  INFORMATION FOR SEQ ID NO: 9:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 126 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;            DESCRIPTION: /desc = "individual isolate"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-573-252-9

  Query Match             65.0%;  Score 13;  DB 4;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAG 16
              |||||||||||||
Db         37 TAGGAGGAACAAG 25


RESULT 26
US-08-248-474-23/c
; Sequence 23, Application US/08248474
; Patent No. 5612471
;  GENERAL INFORMATION:
;    APPLICANT:  McK. BIRD, David
;    APPLICANT:  WILSON, Mark A.
;    TITLE OF INVENTION:  NEMATODE-INDUCE GENES IN TOMATO
;    NUMBER OF SEQUENCES:  114
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/248,474
;      FILING DATE:  25-MAY-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  2307E-535
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum cv 'Rutgers Large
;      ORGANISM:  Red'
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..243
;      OTHER INFORMATION:  /standard_name= "DB# 139"
US-08-248-474-23

  Query Match             65.0%;  Score 13;  DB 1;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAA 15
              |||||||||||||
Db        133 TTAGGAGGAACAA 121
```

```
RESULT 27
US-08-756-849-23/c
; Sequence 23, Application US/08756849
; Patent No. 6093810
;  GENERAL INFORMATION:
;    APPLICANT:  Bird, David McK.
;    APPLICANT:  Wilson, Mark A.
;    TITLE OF INVENTION:  Nematode-Induced Genes in Tomato
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/756,849
;      FILING DATE:  26-NOV-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/248,474
;      FILING DATE:  25-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  02307O-053510US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Lycopersicon esculentum cv 'Rutgers Large Red'
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..243
;      OTHER INFORMATION:  /standard_name= "DB# 139"
US-08-756-849-23

  Query Match             65.0%;  Score 13;  DB 3;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAA 15
              |||||||||||||
Db        133 TTAGGAGGAACAA 121


RESULT 28
US-09-702-705-1692/c
; Sequence 1692, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1692
;   LENGTH: 450
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(450)
;   OTHER INFORMATION: n = A,T,C or G
US-09-702-705-1692

  Query Match             65.0%;  Score 13;  DB 4;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACA 14
              |||||||||||||
Db        235 CTTAGGAGGAACA 223


RESULT 29
US-09-736-457-1692/c
; Sequence 1692, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1692
;   LENGTH: 450
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(450)
;   OTHER INFORMATION: n = A,T,C or G
US-09-736-457-1692

  Query Match             65.0%;  Score 13;  DB 4;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACA 14
              |||||||||||||
Db        235 CTTAGGAGGAACA 223


RESULT 30
US-08-244-205-4/c
; Sequence 4, Application US/08244205
; Patent No. 5952544
;  GENERAL INFORMATION:
;    APPLICANT:  Browse, John, Kinney, Anthony J.,
```

```
;    APPLICANT:  Pierce, John, Wierzbicki, Anna M.,
;    APPLICANT:  Yadav, Narendra S., Perez-Grau, Luis
;    TITLE OF INVENTION:  Fatty Acid Desaturase Genes
;    TITLE OF INVENTION:  from Plants
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. du Pont de Nemours and Company
;      STREET:  1007 Market Street
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh System, 6.0
;      SOFTWARE:  Microsoft Word, 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/244,205
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/804,259
;      FILING DATE:  4 DECEMBER 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Floyd, Linda A.
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  BB-1036-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 992-4929
;      TELEFAX:  (302) 892-7949
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1525 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Arabidopsis thaliana
;    IMMEDIATE SOURCE:
;      CLONE:  pACF2-2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1350
US-08-244-205-4

  Query Match             65.0%;  Score 13;  DB 2;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 GGAGGAACAAGTC 18
              |||||||||||||
Db        814 GGAGGAACAAGTC 802


RESULT 31
PCT-US92-10284-4/c
; Sequence 4, Application PC/TUS9210284
;  GENERAL INFORMATION:
;    APPLICANT:  Browse, John, Kinney, Anthony J.,
;    APPLICANT:  Pierce, John, Wierzbicki, Anna M.,
;    APPLICANT:  Yadav, Narendra S., Perez-Grau, Luis
;    TITLE OF INVENTION:  Fatty Acid Desaturase Genes
;    TITLE OF INVENTION:  from Plants
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. du Pont de Nemours and Company
;      STREET:  1007 Market Street
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh System, 6.0
;      SOFTWARE:  Microsoft Word, 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/10284
;      FILING DATE:  19921203
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/804,259
;      FILING DATE:  4 DECEMBER 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Floyd, Linda A.
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  BB-1036-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 992-4929
;      TELEFAX:  (302) 892-7949
;      TELEX:  835420
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1525 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Arabidopsis thaliana
;    IMMEDIATE SOURCE:
;      CLONE:  pACF2-2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10..1350
PCT-US92-10284-4

  Query Match             65.0%;  Score 13;  DB 5;  Length 1525;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 GGAGGAACAAGTC 18
              |||||||||||||
Db        814 GGAGGAACAAGTC 802


RESULT 32
PCT-US94-01321-9/c
; Sequence 9, Application PC/TUS9401321
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Altered Linolenic and Linoleic Acid Content
;    TITLE OF INVENTION:  in Plants
;    NUMBER OF SEQUENCES:  72
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/01321
;      FILING DATE:. 04-FEB-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/156551
;      FILING DATE:  22-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/014431
;      FILING DATE:  05-FEB-1993
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
```

; LENGTH: 1645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..1465
PCT-US94-01321-9

Query Match 65.0%; Score 13; DB 5; Length 1645;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GGAGGAACAAGTC 18
||||||||||||||
Db 929 GGAGGAACAAGTC 917

RESULT 33
US-09-389-956-79/c
; Sequence 79, Application US/09389956
; Patent No. 6586579
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: PR-Domain Containing Nucleic Acids, Polypeptides,
; TITLE OF INVENTION: Antibodies and Methods
; FILE REFERENCE: P-LJ 3611
; CURRENT APPLICATION NUMBER: US/09/389,956
; CURRENT FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 1970
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1596)
US-09-389-956-79

Query Match 65.0%; Score 13; DB 4; Length 1970;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGT 17
|||||||||||||
Db 139 AGGAGGAACAAGT 127

RESULT 34
US-10-020-079-39/c
; Sequence 39, Application US/10020079
; Patent No. 6579710
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-020-079-39

Query Match 65.0%; Score 13; DB 4; Length 2517;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACA 14
|||||||||||||
Db 1296 CTTAGGAGGAACA 1284

RESULT 35
US-10-020-079-37/c
; Sequence 37, Application US/10020079
; Patent No. 6579710
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 2556
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-020-079-37

Query Match 65.0%; Score 13; DB 4; Length 2556;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACA 14
|||||||||||||
Db 1335 CTTAGGAGGAACA 1323

Search completed: August 15, 2003, 11:00:20
Job time : 36.25 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:57:44 ; Search time 126 Seconds
(without alignments)
428.482 Million cell updates/sec

Title: US-10-074-620-2
Perfect score: 20
Sequence: 1 ccttaggaggaacaagtccc 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20	100.0	20	16	AAQ91012	Primer binding to
2	20	100.0	20	24	ABA00269	EBNA 2 primer, Pos
3	20	100.0	2721	6	AAN50114	DNA sequence encod
4	20	100.0	5931	16	AAT04821	EBV gp350/220 cDNA
5	16	80.0	535	24	ABV96868	Human pancreatic c
6	16	80.0	556	24	ABV99010	Human pancreatic c
7	16	80.0	2376	22	AAH62708	Shrimp white spot
8	16	80.0	6572	22	AAI58931	Human polynucleoti
9	16	80.0	7445	22	ABA09554	Human guanine nucl
10	16	80.0	7445	22	AAK51763	Human polynucleoti
11	16	80.0	7445	22	AAK52747	Human polynucleoti
12	16	80.0	7445	22	AAI60717	Human polynucleoti
13	16	80.0	7843	23	AAS82621	DNA encoding novel
14	16	80.0	9501	25	ACC50083	Breast cancer asso
15	16	80.0	9783	24	ABK94929	Human novel polynu
16	16	80.0	9785	23	ABV22178	Human prostate exp
17	16	80.0	9785	23	ABV28017	Human prostate exp
c 18	16	80.0	180557	24	ABN85750	Human BAC clone RP
19	16	80.0	305107	22	AAH62689	Shrimp white spot
20	15	75.0	123	21	AAC20071	Human secreted pro
c 21	15	75.0	762	24	ABK99951	DNA encoding human
22	15	75.0	1691	24	ABZ66827	Arabidopsis thalia
23	15	75.0	17255	23	ABL29548	Drosophila melanog
24	15	75.0	53226	25	ABQ76896	Human G-protein co
25	14	70.0	125	15	AAQ76831	Human genome fragm
26	14	70.0	273	25	ABX30187	Human GDP-mannose
c 27	14	70.0	300	20	AAX98399	Human cancer cell
28	14	70.0	344	23	ABV06501	Human prostate exp
29	14	70.0	357	21	AAC03189	Human secreted pro
c 30	14	70.0	408	22	AAS38136	Novel human diagno
c 31	14	70.0	409	20	AAX55494	Rice Crm1 homologu
32	14	70.0	442	22	AAS26323	Human cDNA encodin
33	14	70.0	442	25	ABX73664	Human novel polynu
34	14	70.0	467	23	ABV36454	Human prostate exp
c 35	14	70.0	477	22	ABA43976	Human breast cell
c 36	14	70.0	477	22	AAK02723	Human brain expres
c 37	14	70.0	477	22	AAK28164	Human bone marrow
c 38	14	70.0	477	24	ABS02670	Human genome-deriv
39	14	70.0	484	24	ABN95344	Gene #1842 used to
40	14	70.0	484	24	ABL66312	Lung cancer relate
41	14	70.0	494	22	AAK79263	Human immune/haema
42	14	70.0	505	21	AAC35496	Arabidopsis thalia
43	14	70.0	688	23	ABV24146	Human prostate exp
44	14	70.0	764	22	AAI95704	Human neuroblastom
45	14	70.0	798	22	AAS25868	Human cDNA encodin
46	14	70.0	798	25	ABX73209	Human novel polynu
47	14	70.0	813	22	AAH04091	Human cDNA clone (
48	14	70.0	854	21	AAA06579	Human immunogenic
49	14	70.0	854	22	AAS63788	Human prostate cDN
50	14	70.0	854	22	AAH93695	Human prostate-spe
51	14	70.0	854	22	AAH85009	Human prostate-spe
52	14	70.0	854	22	AAH02760	Prostate tumour an
53	14	70.0	854	22	AAF86942	Human P776P invent
54	14	70.0	854	24	ABL95159	Human P776P cDNA s
55	14	70.0	854	25	ACA59596	Prostate cancer th
56	14	70.0	1134	23	ABL27133	Drosophila melanog
c 57	14	70.0	1245	18	AAT78772	DNA encoding Aquif
c 58	14	70.0	1245	25	ABX77244	Aquifex aspartate
59	14	70.0	1541	25	ABX70949	Novel human cDNA s
60	14	70.0	1632	22	AAH13876	Human cDNA sequenc
c 61	14	70.0	1701	21	AAC44045	Zea mays DNA fragm
62	14	70.0	1763	22	AAH45659	Human protease reg
c 63	14	70.0	1821	21	AAC51101	Arabidopsis thalia
64	14	70.0	3134	23	ABL27132	Drosophila melanog
65	14	70.0	3522	23	ABX71253	Human brain-derive
66	14	70.0	3529	22	AAH76684	Human transcriptio
c 67	14	70.0	4619	23	ABX71386	Human testes-deriv
c 68	14	70.0	4809	22	AAS63924	Human prostate cDN
c 69	14	70.0	4809	22	AAH93831	Human prostate-spe
c 70	14	70.0	4809	24	ABL95295	Human P776P cDNA s
c 71	14	70.0	4809	25	ACA59732	Prostate cancer th
c 72	14	70.0	6250	22	AAL36600	Human musculoskele
c 73	14	70.0	6250	25	ABX59588	cDNA encoding nove
c 74	14	70.0	6251	22	AAL36603	Human musculoskele
c 75	14	70.0	6251	25	ABX59591	cDNA encoding nove
76	14	70.0	9636	22	AAK90449	Human digestive sy
77	14	70.0	11446	22	AAK90464	Human digestive sy
78	14	70.0	16607	22	ABA18349	Human nervous syst
c 79	14	70.0	16607	22	AAL04678	Human reproductive
c 80	14	70.0	16607	23	ABL97585	Human testicular a
81	14	70.0	16738	22	AAK70864	Human immune/haema

c 82	14	70.0	24025	17	AAT17455	Mutated BRCA1 geno
c 83	14	70.0	24025	17	AAT17515	Mutated BRCA1 geno
c 84	14	70.0	24026	17	AAT32612	BRCA1, human breas
c 85	14	70.0	24026	17	AAT17512	Mutated BRCA1 geno
c 86	14	70.0	24026	17	AAT17513	Mutated BRCA1 geno
c 87	14	70.0	24026	17	AAT17514	Mutated BRCA1 geno
c 88	14	70.0	24026	17	AAT17516	Mutated BRCA1 geno
c 89	14	70.0	24026	17	AAT17517	Mutated BRCA1 geno
c 90	14	70.0	24026	17	AAT17518	Mutated BRCA1 geno
c 91	14	70.0	24026	17	AAT17519	Mutated BRCA1 geno
c 92	14	70.0	24026	17	AAT17521	Mutated BRCA1 geno
c 93	14	70.0	24026	17	AAT17522	Mutated BRCA1 geno
c 94	14	70.0	24026	17	AAT17523	Mutated BRCA1 geno
c 95	14	70.0	24026	17	AAT17524	Mutated BRCA1 geno
c 96	14	70.0	24026	17	AAT17526	Mutated BRCA1 geno
c 97	14	70.0	24026	17	AAT17527	Mutated BRCA1 geno
c 98	14	70.0	24026	17	AAT17528	Mutated BRCA1 geno
c 99	14	70.0	24026	17	AAT17529	Mutated BRCA1 geno
c 100	14	70.0	24026	17	AAT17530	BRCA1 genomic sequ
c 101	14	70.0	24026	17	AAT18325	BRCA1, human breas
c 102	14	70.0	24029	17	AAT17520	Mutated BRCA1 geno
c 103	14	70.0	24031	17	AAT17525	Mutated BRCA1 geno
104	14	70.0	25574	22	AAL05619	Human reproductive
105	14	70.0	25574	22	AAK79671	Human immune/haema
106	14	70.0	25574	22	AAK83760	Human immune/haema
107	14	70.0	25576	22	AAL05618	Human reproductive
108	14	70.0	25576	22	AAK79669	Human immune/haema
109	14	70.0	25576	22	AAK83758	Human immune/haema
110	14	70.0	25576	22	AAK85305	Human immune/haema
c 111	14	70.0	77425	24	ABK83502	Human cDNA differe
c 112	14	70.0	567571	25	AAD53224	Human chromosome 3
113	14	70.0	1691080	24	ABX08336	Human phosphodiest
114	13	65.0	114	19	AAV44485	tRNA-Lys3 pseudoge
c 115	13	65.0	126	18	AAT68500	Clone (individual)
c 116	13	65.0	126	18	AAT68881	Clone (individual)
c 117	13	65.0	242	22	ABA11686	Human nervous syst
c 118	13	65.0	243	17	AAT09746	Tomato genomic DNA
119	13	65.0	286	24	ABL73501	Corn tassel-derive
c 120	13	65.0	294	24	ABL75621	Corn tassel-derive

ALIGNMENTS

RESULT 1
AAQ91012
ID AAQ91012 standard; DNA; 20 BP.
XX
AC AAQ91012;
XX
DT 01-FEB-1996 (first entry)
XX
DE Primer binding to 3' end of EBV nuc antigen gene.
XX
KW Primer; PCR; amplification; DNA polymerase; exonuclease; Pfu; Taq;
KW Klenow fragment; T4; T7; Deep Vent; synthesis; mismatch; human; antibody;
KW heavy chain variable region; ss.
XX
OS Synthetic.
XX
PN WO9516028-A1.
XX
PD 15-JUN-1995.
XX
PF 07-DEC-1994; 94WO-US14065.
XX
PR 16-FEB-1994; 94US-0197791.
PR 08-DEC-1993; 93US-0164290.
XX
PA (STRA-) STRATAGENE.
XX
PI Mullinax RL, Sorge JA;
XX
DR WPI; 1995-224316/29.
XX
PT Compsn. useful for polynucleotide synthesis and cyclical
PT amplification - comprising a mixt. contg. an enzyme with 3'-5'
PT exo-nuclease activity and a DNA polymerase with less 3'-5'
PT exo-nuclease activity than the enzyme
XX
PS Examples; Page 35; 66pp; English.
XX
CC Primers AAQ90984-Q91028 are examples of primers for testing a novel
CC composition for polynucleotide synthesis comprising a DNA polymerase
CC with high 3'-5' exonuclease activity in conjunction with a DNA polymerase
CC with less 3'-5' exonuclease activity, pref. Pfu and Taq DNA polymerases
CC respectively. Other DNA polymerases containing high 3'-5' exonuclease
CC activity include E.coli DNA polymerase I, Klenow fragment, T4, T7, Vent
CC or Deep Vent DNA polymerases. The use of a DNA polymerase with high
CC 3'-5' exonuclease activity is designed to overcome the inability of DNA
CC polymerases with low 3'-5' exonuclease activities to initiate synthesis
CC from primers containing 3' terminal mismatches, e.g. due to errors
CC introduced during a PCR cycle.
CC This primer binds to a region at the 5' end of the Epstein-Barr virus
CC nuc antigen gene.
XX
SQ Sequence 20 BP; 6 A; 6 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 16; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCTTAGGAGGAACAAGTCCC 20
||||||||||||||||||||
Db 1 CCTTAGGAGGAACAAGTCCC 20

RESULT 2
ABA00269
ID ABA00269 standard; DNA; 20 BP.
XX
AC ABA00269;
XX
DT 29-NOV-2002 (first entry)
XX
DE EBNA 2 primer, Position 90249:90269.
XX
KW Primer; amplify; PCR; probe; detection; Epstein-Barr virus; EBV; ss.
XX
OS Epstein-barr virus.
XX
PN WO200264842-A2.
XX
PD 22-AUG-2002.
XX
PF 13-FEB-2002; 2002WO-US04339.
XX
PR 13-FEB-2001; 2001US-268439P.
XX
PA (CHIL-) CHILDRENS HOSPITAL RES FOUND.
XX
PI Witte DP, Groen PA;
XX
DR WPI; 2002-667015/71.
XX
PT New compositions comprising nucleic acid sequences which specifically
PT hybridizes to Epstein-Barr virus (EBV) nucleic acid, for detecting EBV
PT in clinical specimens to determine patients at high risk of to
PT developing EBV infections -
XX
PS Claim 1; Page 44; 59pp; English.
XX
CC The sequences given in ABA00268-75 are primers and probes which were
CC used in the compositions of the invention for the detection of
CC Epstein-Barr virus (EBV). The compositions comprise at least one
CC purified and isolated oligonucleotide consisting of a nucleic acid

```
CC   sequence which complements and specifically hybridizes to EBV nucleic
CC   acid. The oligonucleotide sequences and compositions comprising them
CC   are useful for detecting EBV in clinical specimens to determine
CC   patients who are at high risk to develop serious and costly medical
CC   complications, and allow for better clinical management of these
CC   patients by earlier recognition of their infection status. The
CC   oligonucleotide sequences may also be used to amplify EBV DNA
CC   sequences. The use of the oligonucleotide sequences in the assay for
CC   detecting EBV has a broad dynamic range of detection from less than
CC   10-100000000 copies of EBV DNA, is less labour intensive requiring only
CC   one reaction tube for the EBV determination, highly sensitive, accurate
CC   and has a rapid turn around time with assays that are completed,
CC   including amplification, probe specific hybridization, and calculation
CC   of copy number in less than 1 hour. The method may be adapted to
CC   automated systems.
XX
SQ   Sequence 20 BP; 6 A; 6 C; 5 G; 3 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 24;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.078;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db          1 CCTTAGGAGGAACAAGTCCC 20


RESULT 3
AAN50114
ID   AAN50114 standard; DNA; 2721 BP.
XX
AC   AAN50114;
XX
DT   25-MAR-2003  (updated)
DT   17-OCT-1991  (first entry)
XX
DE   DNA sequence encoding Epstein-Barr virus (EBV) outer surface protein.
XX
KW   Epstein-Barr virus; antigen; vaccine; ss.
XX
OS   Epstein-Barr virus.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1..2721
FT                   /*tag=  a
FT                   /label= EBV surface protein antigen
XX
PN   EP151079-A.
XX
PD   07-AUG-1985.
XX
PF   28-JAN-1985;    85EP-0400141.
XX
PR   23-JUL-1984;    84US-0633558.
PR   30-JAN-1984;    84US-0575352.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Kleff E,  Tanner J,  Hummel M,  Belsel C;
XX
DR   WPI; 1985-191978/32.
DR   P-PSDB; AAP50073.
XX
PT   New fragment of Epstein-Barr Virus DNA - useful in vector to
PT   express polypeptide for use in prepn. of vaccine against the
PT   virus and for use in diagnosis.
XX
PS   Claim 1; Page 21-23; 26pp; English.
XX
CC   The sequence encodes an outer surface viral protein of EBV, used
CC   to generate antibodies reacting with the surface proteins of
CC   EBV-infected cells, and in the preparation of a vaccine against EBV.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2721 BP; 762 A; 876 C; 557 G; 526 T; 0 other;

  Query Match             100.0%;  Score 20;  DB 6;  Length 2721;
  Best Local Similarity   100.0%;  Pred. No. 0.058;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       1886 CCTTAGGAGGAACAAGTCCC 1905


RESULT 4
AAT04821
ID   AAT04821 standard; cDNA; 5931 BP.
XX
AC   AAT04821;
XX
DT   18-JAN-1996  (first entry)
XX
DE   EBV gp350/220 cDNA.
XX
KW   EBV; gp350; gp220; gp350/gp220; non-splicing variant; vaccine; ds.
XX
OS   Epstein-Barr virus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1014..3737
FT                   /*tag=  a
FT   sig_peptide     1014..1067
FT                   /*tag=  b
FT   mat_peptide     1068..3734
FT                   /*tag=  c
FT   misc_feature    2514..2515
FT                   /*tag=  d
FT                   /function= splice donor site
FT                   /note= "bases 2513-2517 (AAGT) are replaced by
FT                         GTCA in the non-splicing variant"
FT   misc_feature    3105..3106
FT                   /*tag=  e
FT                   /function= splice acceptor site
FT                   /note= "bases 3104-3107 (AGGT) are replaced by
FT                         TGGA in the non-splicing variant"
FT   polyA_signal    3742..3747
FT                   /*tag=  f
XX
PN   WO9528488-A1.
XX
PD   26-OCT-1995.
XX
PF   13-APR-1995;    95WO-US04611.
XX
PR   18-APR-1994;    94US-0229291.
XX
PA   (AVIR-) AVIRON.
XX
PI   Jackman WT,  Spaete R;
XX
DR   WPI; 1995-373802/48.
DR   P-PSDB; AAR80144.
XX
PT   New DNA encoding a homogeneous gp350 protein - can be used for
PT   preventing and treating Epstein-Barr virus-related diseases or
PT   conditions
XX
PS   Claim 2; Fig.1; 61pp; English.
XX
CC   The donor and acceptor splice sites of the EBV gene encoding gp350/
CC   220 are mutated by replacement of native nucleotides by non-native
CC   nucleotides, without altering the encoded amino acid sequence,
CC   resulting in elimination of gp220 prodn.  Recombinant homogeneous
CC   gp350, useful in vaccines, is expressed in mammalian or insect cell
CC   hosts.
```

```
XX
SQ   Sequence 5931 BP; 1453 A; 1782 C; 1437 G; 1259 T; 0 other;

  Query Match            100.0%;  Score 20;  DB 16;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 0.055;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 5
ABV96868
ID   ABV96868 standard; cDNA; 535 BP.
XX
AC   ABV96868;
XX
DT   14-JAN-2003  (first entry)
XX
DE   Human pancreatic cancer expressed cDNA SEQ ID NO 2276.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US02781.
XX
PR   30-JAN-2001; 2001US-265305P.
PR   31-JAN-2001; 2001US-265682P.
PR   09-FEB-2001; 2001US-267568P.
PR   21-MAR-2001; 2001US-278651P.
PR   28-APR-2001; 2001US-287112P.
PR   16-MAY-2001; 2001US-291631P.
PR   12-JUL-2001; 2001US-305484P.
PR   20-AUG-2001; 2001US-313999P.
PR   27-NOV-2001; 2001US-333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR,  Kalos MD,  Lodes MJ,  Persing DH,  Hepler WT,  Jiang Y;
XX
DR   WPI; 2002-627435/67.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful
PT   for diagnosing, preventing and/or treating cancer, particularly
PT   pancreatic cancer  -
XX
PS   Claim 1; SEQ ID NO 2276; 300pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145);
CC   (b) complements of (a); (c) sequences consisting of at least 20
CC   contiguous residues of (a); (d) sequences that hybridize to (a), under
CC   moderately stringent conditions; (e) sequences having at least 75% or 90%
CC   identity to (a); or (f) degenerate variants of (a). Polypeptides
CC   (ABP68596-ABP68637) encoded by (I) and oligonucleotide can be used to
CC   detect cancer in a patient and compositions comprising polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations and
CC   antigen presenting cells expressing the polypeptide are useful in
CC   treating pancreatic cancer and stimulating an immune response. The
CC   polynucleotides can be used as probes or primers for nucleic acid
CC   hybridisation, in the design and preparation of ribozyme molecules for
CC   inhibiting expression of the tumour polypeptides and proteins in the
CC   tumour cells, in vaccines and for gene therapy.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 535 BP; 171 A; 112 C; 104 G; 148 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 24;  Length 535;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         87 CTTAGGAGGAACAAGT 102


RESULT 6
ABV99010
ID   ABV99010 standard; cDNA; 556 BP.
XX
AC   ABV99010;
XX
DT   14-JAN-2003  (first entry)
XX
DE   Human pancreatic cancer expressed cDNA SEQ ID NO 4418.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US02781.
XX
PR   30-JAN-2001; 2001US-265305P.
PR   31-JAN-2001; 2001US-265682P.
PR   09-FEB-2001; 2001US-267568P.
PR   21-MAR-2001; 2001US-278651P.
PR   28-APR-2001; 2001US-287112P.
PR   16-MAY-2001; 2001US-291631P.
PR   12-JUL-2001; 2001US-305484P.
PR   20-AUG-2001; 2001US-313999P.
PR   27-NOV-2001; 2001US-333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR,  Kalos MD,  Lodes MJ,  Persing DH,  Hepler WT,  Jiang Y;
XX
DR   WPI; 2002-627435/67.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful
PT   for diagnosing, preventing and/or treating cancer, particularly
PT   pancreatic cancer  -
XX
PS   Claim 1; SEQ ID NO 4418; 300pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145);
CC   (b) complements of (a); (c) sequences consisting of at least 20
CC   contiguous residues of (a); (d) sequences that hybridize to (a), under
CC   moderately stringent conditions; (e) sequences having at least 75% or 90%
CC   identity to (a); or (f) degenerate variants of (a). Polypeptides
CC   (ABP68596-ABP68637) encoded by (I) and oligonucleotide can be used to
CC   detect cancer in a patient and compositions comprising polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations and
CC   antigen presenting cells expressing the polypeptide are useful in
CC   treating pancreatic cancer and stimulating an immune response. The
CC   polynucleotides can be used as probes or primers for nucleic acid
CC   hybridisation, in the design and preparation of ribozyme molecules for
CC   inhibiting expression of the tumour polypeptides and proteins in the
CC   tumour cells, in vaccines and for gene therapy.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence 556 BP; 173 A; 117 C; 114 G; 152 T; 0 other;

  Query Match            80.0%;  Score 16;  DB 24;  Length 556;
  Best Local Similarity  100.0%;  Pred. No. 10;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        108 CTTAGGAGGAACAAGT 123


RESULT 7
AAH62708
ID   AAH62708 standard; DNA; 2376 BP.
XX
AC   AAH62708;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Shrimp white spot Bacilliform virus (WSBV) gene 19.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp; ds.
XX
OS   White spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US28888.
XX
PR   24-NOV-1999;   99CN-0124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   P-PSDB; AAG84928.
XX
PT   Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection -
XX
PS   Claim 4; Figure 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp.
XX
SQ   Sequence 2376 BP; 787 A; 467 C; 526 G; 596 T; 0 other;

  Query Match            80.0%;  Score 16;  DB 22;  Length 2376;
  Best Local Similarity  100.0%;  Pred. No. 9.3;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        793 AGGAGGAACAAGTCCC 808


RESULT 8
AAI58931
ID   AAI58931 standard; cDNA; 6572 BP.
XX
AC   AAI58931;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1134.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39775.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 1134; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 6572 BP; 1929 A; 1256 C; 1418 G; 1969 T; 0 other;

  Query Match            80.0%;  Score 16;  DB 22;  Length 6572;
  Best Local Similarity  100.0%;  Pred. No. 8.8;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
```

```
Db       4524 CTTAGGAGGAACAAGT 4539


RESULT 9
ABA09554
ID   ABA09554 standard; cDNA; 7445 BP.
XX
AC   ABA09554;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human guanine nucleotide exchange factor homologue cDNA, SEQ ID NO:1330.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytcstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   P-PSDB; ABB12310.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 1; Page 992-994; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a cDNA encoding a
CC   novel human polypeptide of the invention.
XX
SQ   Sequence 7445 BP; 2177 A; 1404 C; 1582 G; 2282 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 22;  Length 7445;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db       4492 CTTAGGAGGAACAAGT 4507


RESULT 10
AAK51763
ID   AAK51763 standard; cDNA; 7445 BP.
XX
AC   AAK51763;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 308.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM78630.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 1279-1284; 6221pp; English.
XX
```

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 7445 BP; 2179 A; 1403 C; 1584 G; 2279 T; 0 other;

Query Match 80.0%; Score 16; DB 22; Length 7445;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 4481 CTTAGGAGGAACAAGT 4496

RESULT 11
AAK52747
ID AAK52747 standard; cDNA; 7445 BP.
XX
AC AAK52747;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2276.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79614.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 4614-4616; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 7445 BP; 2177 A; 1404 C; 1582 G; 2282 T; 0 other;

Query Match 80.0%; Score 16; DB 22; Length 7445;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 4492 CTTAGGAGGAACAAGT 4507

RESULT 12
AAI60717
ID AAI60717 standard; cDNA; 7445 BP.
XX
AC AAI60717;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4706.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41561.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4706; 10078pp; English.
XX

```
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 7445 BP; 2177 A; 1404 C; 1582 G; 2282 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 22;  Length 7445;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db       4492 CTTAGGAGGAACAAGT 4507


RESULT 13
AAS82621
ID   AAS82621 standard; cDNA; 7843 BP.
XX
AC   AAS82621;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #18425.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG18434.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 18425; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7843 BP; 2262 A; 1492 C; 1682 G; 2403 T; 4 other;

  Query Match             80.0%;  Score 16;  DB 23;  Length 7843;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db       4761 CTTAGGAGGAACAAGT 4776


RESULT 14
ACC50083
ID   ACC50083 standard; cDNA; 9501 BP.
XX
AC   ACC50083;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Breast cancer associated cDNA sequence SEQ ID NO:14.
XX
KW   Human; breast cancer; cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003004989-A2.
XX
PD   16-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-US19669.
XX
PR   21-JUN-2001; 2001US-299887P.
PR   27-JUN-2001; 2001US-301572P.
PR   18-JUL-2001; 2001US-306501P.
PR   25-SEP-2001; 2001US-325002P.
PR   05-MAR-2002; 2002US-362585P.
PR   14-MAY-2002; 2002US-380391P.
XX
PA   (MILL-) MILLENIUM PHARM INC.
XX
PI   Lillie J,  Gannavarapu M,  Glatt K,  Hoersh S,  Kamatkar S,  Mertens M;
PI   Monahan JE,  Myer V,  Wang Y,  Xu Y,  Zhao X,  Meyers RE,  Bast RC;
PI   Hortobagyi GN,  Pusztai L,  Meric F,  Sahin A,  Mills GB;
XX
DR   WPI; 2003-210381/20.
DR   P-PSDB; ABR47392.
XX
PT   Breast cancer diagnosis or treatment by comparing the level of
PT   expression of a marker in a patient sample with that in the control
PT   non-breast cancer sample  -
XX
PS   Claim 1; SEQ ID 14; 128pp; English.
XX
CC   The present invention describes a method for assessing whether a patient
CC   is afflicted with breast cancer. The method comprises comparing the level
CC   of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC   ABR47386 to ABR47632) in a patient sample and the normal level of
CC   expression of the marker in a control non-breast cancer sample, where a
```

CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences
CC from the present invention have cytostatic activities and can be used in
CC gene therapy. The method is useful for diagnosing and treating breast
CC cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9501 BP; 2816 A; 1872 C; 2078 G; 2735 T; 0 other;

Query Match 80.0%; Score 16; DB 25; Length 9501;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 6556 CTTAGGAGGAACAAGT 6571

RESULT 15
ABK94929
ID ABK94929 standard; cDNA; 9783 BP.
XX
AC ABK94929;
XX
DT 30-AUG-2002 (first entry)
XX
DE Human novel polynucleotide #40.
XX
KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.
XX
OS Homo sapiens.
XX
PN WO200244340-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US47004.
XX
PR 30-NOV-2000; 2000US-0028952.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX
DR WPI; 2002-508509/54.
DR P-PSDB; ABG66705.
XX
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX
PS Claim 1; Page 425-434; 672pp; English.
XX
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human
CC novel polynucleotides of the invention.
XX
SQ Sequence 9783 BP; 2881 A; 1927 C; 2156 G; 2819 T; 0 other;

Query Match 80.0%; Score 16; DB 24; Length 9783;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 6826 CTTAGGAGGAACAAGT 6841

RESULT 16
ABV22178
ID ABV22178 standard; cDNA; 9785 BP.
XX
AC ABV22178;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 22169.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 3813-3814; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 9785 BP; 2870 A; 1936 C; 2160 G; 2819 T; 0 other;

Query Match 80.0%; Score 16; DB 23; Length 9785;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 6835 CTTAGGAGGAACAAGT 6850

RESULT 17
ABV28017
ID ABV28017 standard; cDNA; 9785 BP.
XX
AC ABV28017;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 28008.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2C0160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 5783-5785; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 9785 BP; 2870 A; 1936 C; 2160 G; 2819 T; 0 other;

Query Match 80.0%; Score 16; DB 23; Length 9785;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 6835 CTTAGGAGGAACAAGT 6850

RESULT 18
ABN85750/c
ID ABN85750 standard; DNA; 180557 BP.
XX
AC ABN85750;
XX
DT 14-OCT-2002 (first entry)
XX
DE Human BAC clone RP11-334G22 SEQ ID NO 6.
XX
KW Human; Can 1; antiinfertility; gynaecological; infertility;
KW premature ovarian failure; menopause; Sertoli Cell only syndrome;
KW BAC clone RP11-334G22; GenBank reference AC007250; ds.
XX
OS Homo sapiens.
XX
PN US2002119929-A1.
XX
PD 29-AUG-2002.
XX
PF 02-NOV-2001; 2001US-0003806.
XX
PR 03-NOV-2000; 2000US-245872P.
XX
PA (BISH/) BISHOP C E.
PA (AGOU/) AGOULNIK A I.
PA (ZHUQ/) ZHU Q.
XX
PI Bishop CE, Agoulnik AI, Zhu Q;
XX
DR WPI; 2002-618953/66.
XX
PT A nucleic acid molecule (I) encoding a Can 1 polypeptide used in
PT treating infertility -
XX
PS Disclosure; Page -; 45pp; English.
XX
CC The invention relates to a nucleic acid molecule (I) encoding a Can 1
CC polypeptide. The Can 1 nucleic acid molecule is used to diagnose or treat
CC infertility or premature ovarian failure or Sertoli Cell only syndrome
CC in a mammal. The present sequence is that of a human Can 1 encoding
CC BAC clone RP11-334G22 of the invention.
CC Note: The present sequence is not given in the printed specification but
CC was obtained through the GenBank reference AC007250.
XX
SQ Sequence 180557 BP; 53238 A; 32016 C; 33894 G; 61409 T; 0 other;

Query Match 80.0%; Score 16; DB 24; Length 180557;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 124080 AGGAGGAACAAGTCCC 124065

RESULT 19
AAH62689

ID AAH62689 standard; DNA; 305107 BP.
XX
AC AAH62689;
XX
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) genomic sequence.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp; ds.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.
XX
PD 31-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US28888.
XX
PR 24-NOV-1999; 99CN-0124717.
XX
PA (PENY-) PE CORP NY.
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA (SINO-) SINOGENOMAX CO LTD.
XX
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX
DR WPI; 2001-355877/37.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection -
XX
PS Disclosure; Figure 2; 626pp; English.
XX
CC The invention provides the primary nucleotide sequence of the WSBV genome
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp.
XX
SQ Sequence 305107 BP; 92042 A; 62482 C; 62635 G; 87948 T; 0 other;

Query Match 80.0%; Score 16; DB 22; Length 305107;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 155349 AGGAGGAACAAGTCCC 155364

RESULT 20
AAC20071
ID AAC20071 standard; cDNA; 123 BP.
XX
AC AAC20071;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 24146.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 24146; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 123 BP; 48 A; 15 C; 22 G; 33 T; 5 other;

Query Match 75.0%; Score 15; DB 21; Length 123;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTAGGAGGAACAAGT 17
|||||||||||||||
Db 19 TTAGGAGGAACAAGT 33

RESULT 21
ABK99951/c
ID ABK99951 standard; DNA; 762 BP.
XX
AC ABK99951;
XX
DT 21-OCT-2002 (first entry)
XX
DE DNA encoding human secreted protein SCEP-34.
XX
KW Secreted protein; SCEP; human; cell proliferative disorder; cancer;
KW keratosis; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
KW psoriasis; autoimmune disorder; inflammatory disorder; AIDS; arthritis;
KW acquired immunodeficiency syndrome; adult respiratory distress syndrome;
KW Addison's disease; allergy; asthma; osteoporosis; autoimmune thyroiditis;
KW Crohn's disease; dermatitis; diabetes; Graves' disease; haemodialysis;
KW glomerulonephritis; scleroderma; systemic lupus erythematosus; uveitis;
KW systemic sclerosis; ulcerative colitis; infection; trauma; Pick disease;
KW cardiovascular disorder; congestive heart failure; angina; epilepsy;
KW hypertensive heart disease; neurological disorder; Alzheimer's disesae;
KW Parkinson's disease; amyotrophic lateral sclerosis; stroke; dementia;
KW Huntington's disease; multiple sclerosis; neuromuscular disorder;
KW metabolic disorder; endocrine disorder; toxic myopathy; mental disorder;
KW schizophrenic disorder; developmental disorder; anaemia; epilepsy;
KW hypothyroidism; glaucoma; sensorineural hearing loss; cataract;
KW transgenic animal; gene; ds.
XX

OS Homo sapiens.
XX
PN WO200248337-A2.
XX
PD 20-JUN-2002.
XX
PF 12-DEC-2001; 2001WO-US48517.
XX
PR 13-DEC-2000; 2000US-255639P.
PR 21-DEC-2000; 2000US-257852P.
PR 05-JAN-2001; 2001US-260105P.
PR 18-JAN-2001; 2001US-262932P.
PR 18-JAN-2001; 2001US-263096P.
PR 19-JAN-2001; 2001US-263090P.
PR 02-FEB-2001; 2001US-265926P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Griffin JA, Yao MG, Duggan BM, Yue H, Ding L, Lal PG, Lee EA;
PI Ramkumar J, Thangavelu K, Xu Y, Lee S, Tang YT, Nguyen DB;
PI Warren BA, Honchell CD, Gietzen KJ, Baughn MR, Gandhi AR;
PI Arvizu C, Walia NK, Lu Y, Elliott VW, Lu DAM, Hafalia AJA;
PI Azimzai Y, Khan FA, Tran UK;
XX
DR WPI; 2002-583509/62.
DR P-PSDB; ABG69654.
XX
PT Novel human secreted proteins and polynucleotides for diagnosing,
PT preventing or treating cell proliferative, autoimmune/inflammatory,
PT cardiovascular, neurological and developmental disorders -
XX
PS Claim 5; Page 221; 234pp; English.
XX
CC The invention describes an isolated polypeptide chosen from secreted
CC proteins (I), SECP 1-54. (I) and the polynucleotide encoding it (II) are
CC useful for screening a compound for effectiveness as an agonist or
CC antagonist of (I) or compound that alters expression of (II). (I), the
CC identified agonist and antagonist are useful for treating a disease or
CC condition associated altered expression of functional SECP in a patient.
CC An antibody specific to (I) is useful for detecting the presence of (I),
CC purifying (I) from a sample and for diagnosing a condition or disease
CC associated with expression of SECP in a subject or in a biological
CC sample. (I) and (II) and modulators of (I) are useful for diagnosis,
CC treatment and prevention of cell proliferative disorders (e.g. cancer,
CC keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and
CC psoriasis), autoimmune/inflammatory disorders (e.g. acquired
CC immunodeficiency syndrome (AIDS), adult respiratory distress syndrome,
CC Addison's disease, allergies, asthma, osteoporosis, autoimmune
CC thyroiditis, Crohn's disease, dermatitis, diabetes, Graves' disease,
CC glomerulonephritis, arthritis, scleroderma, systemic lupus erythematosus,
CC systemic sclerosis, ulcerative colitis, haemodialysis, uveitis; viral,
CC bacterial, fungal, parasitic, protozoal, helminthic infections and
CC trauma), cardiovascular disorders (e.g. congestive heart failure, angina,
CC hypertensive heart disease), neurological disorders (e.g. Alzheimer's and
CC Pick disease, Parkinson disease, amyotrophic lateral sclerosis, epilepsy,
CC stroke, Huntington's disease, multiple sclerosis, dementia, neuromuscular
CC disorders, metabolic, endocrine and toxic myopathies, mental disorders,
CC schizophrenic disorders, and developmental disorders (e.g. anaemia,
CC epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and
CC cataract). (II) is useful for creating transgenic animals to model human
CC disease and to detect and quantify gene expression in biopsied tissues in
CC which expression of SECP is correlated with disease. This sequence
CC encodes a human secreted protein (SCEP).
XX
SQ Sequence 762 BP; 174 A; 210 C; 164 G; 214 T; 0 other;

Query Match 75.0%; Score 15; DB 24; Length 762;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCC 19
||||||||||||||||
Db 555 AGGAGGAACAAGTCC 541

RESULT 22
ABZ66827
ID ABZ66827 standard; DNA; 1691 BP.
XX
AC ABZ66827;
XX
DT 24-MAR-2003 (first entry)
XX
DE Arabidopsis thaliana polynucleotide SEQ ID NO 23.
XX
KW Arabidopsis thaliana; herbicidal; plant; growth regulator; transgenic;
KW herbicide-resistant; gene; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200266660-A2.
XX
PD 29-AUG-2002.
XX
PF 13-FEB-2002; 2002WO-EP01466.
XX
PR 16-FEB-2001; 2001DE-1007843.
PR 23-MAY-2001; 2001DE-1025537.
XX
PA (META-) METANOMICS GMBH & CO KGAA.
XX
PI Plesch G, Blau A, Daeschner K, Klein M;
XX
DR WPI; 2002-674953/72.
XX
PT Identifying herbicides and plant growth regulators, from ability to
PT inhibit specific genes, also use of these genes to prepare
PT herbicide-resistant transgenic plants -
XX
PS Disclosure; Page 154-155; 224pp; German.
XX
CC The invention relates to identifying compounds (A) with herbicidal
CC activity from their ability to reduce or block expression or activity of
CC gene products of specific nucleic acid sequences (I) or amino acid
CC sequences (II) encoded by (I). (A) and antagonists of proteins encoded
CC by (I) are useful as herbicides and plant growth regulators.
CC Overexpression of (I) in plants is used to produce herbicide-resistant
CC plants and (I) may also be mutated to identify altered sequences that
CC encode proteins resistant to herbicides. The present sequence is that of
CC a polynucleotide of the invention.
XX
SQ Sequence 1691 BP; 483 A; 309 C; 434 G; 465 T; 0 other;

Query Match 75.0%; Score 15; DB 24; Length 1691;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TTAGGAGGAACAAGT 17
|||||||||||||||
Db 830 TTAGGAGGAACAAGT 844

RESULT 23
ABL29548
ID ABL29548 standard; DNA; 17255 BP.
XX
AC ABL29548;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40117.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.

```
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 40117; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 17255 BP; 5336 A; 3353 C; 3445 G; 5121 T; 0 other;

  Query Match             75.0%;  Score 15;  DB 23;  Length 17255;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCC 19
              |||||||||||||||
Db      10781 AGGAGGAACAAGTCC 10795


RESULT 24
ABQ76896
ID   ABQ76896 standard; DNA; 53226 BP.
XX
AC   ABQ76896;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human G-protein coupled receptor DNA SEQ ID 3.
XX
KW   G-protein coupled receptor; secretin receptor subfamily; human; SNP;
KW   GPCR; protease; Parkinson's disease; gene; chromosome X;
KW   single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3000..50651
FT                   /*tag=  a
FT                   /product= "GPCR"
FT                   /note "this coding sequence is interrupted by
FT                         13 introns"
FT   variation       replace (1746,c)
FT                   /*tag=  b
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (1755,g)
FT                   /*tag=  c
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (1961,g)
FT                   /*tag=  d
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            3000..3088
FT                   /*tag=  e
FT                   /number= 1
FT   intron          3089..3874
FT                   /*tag=  f
FT                   /number= 1
FT   exon            3875..4038
FT                   /*tag=  g
FT                   /number= 2
FT   intron          4039..6037
FT                   /*tag=  h
FT                   /number= 2
FT   variation       replace (5411,g)
FT                   /*tag=  i
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (5760,a)
FT                   /*tag=  j
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            6038..6170
FT                   /*tag=  k
FT                   /number= 3
FT   intron          6171..8059
FT                   /*tag=  l
FT                   /number= 3
FT   exon            8060..8178
FT                   /*tag=  m
FT                   /number= 4
FT   intron          8179..15910
FT                   /*tag=  n
FT                   /number= 4
FT   variation       replace (11390,c)
FT                   /*tag=  o
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (11390,t)
FT                   /*tag=  p
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            15911..16127
FT                   /*tag=  q
FT                   /number= 5
FT   intron          16128..17484
FT                   /*tag=  r
FT                   /number= 5
FT   variation       replace (16988,a)
FT                   /*tag=  s
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            17485..17647
FT                   /*tag=  t
FT                   /number= 6
FT   intron          17648..32332
FT                   /*tag=  u
FT                   /number= 6
FT   variation       replace (18361,g)
FT                   /*tag=  v
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (19769,c)
FT                   /*tag=  w
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (22910,t)
FT                   /*tag=  x
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (22935,t)
FT                   /*tag=  y
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24206,g)
FT                   /*tag=  z
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24774,a)
FT                   /*tag=  aa
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24869,c)
```

```
FT                   /*tag=  ab
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (25768,a)
FT                   /*tag=  ac
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (26697,t)
FT                   /*tag=  ad
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (26697,c)
FT                   /*tag=  ae
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (28359,c)
FT                   /*tag=  af
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (28470..28471,t)
FT                   /*tag=  ag
FT                   /note= "a single nucleotide polymorphism (SNP) can
FT                          result in a deletion at this position"
FT   variation       replace (29781,g)
FT                   /*tag=  ah
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (30182,a)
FT                   /*tag=  ai
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (31772,t)
FT                   /*tag=  aj
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (31936,g)
FT                   /*tag=  ak
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            32333..32467
FT                   /*tag=  al
FT                   /number= 7
FT   intron          32468..36825
FT                   /*tag=  am
FT                   /number= 7
FT   exon            36826..36948
FT                   /*tag=  an
FT                   /product= 8
FT   intron          36949..38129
FT                   /*tag=  ao
FT                   /number= 8
FT   exon            38130..38175
FT                   /*tag=  ap
FT                   /number= 9
FT   intron          38176..39984
FT                   /*tag=  aq
FT                   /number= 9
FT   exon            39985..40088
FT                   /*tag=  ar
FT                   /number= 10
FT   intron          40089..42455
FT                   /*tag=  as
FT                   /number= 10
FT   exon            42456..42577
FT                   /*tag=  at
FT                   /number= 11
FT   intron          42578..44422
FT                   /*tag=  au
FT                   /number= 11
FT   variation       replace (42767..42767,c)
FT                   /*tag=  av
FT                   /note= "a single nucleotide polymorphism (SNP) can
FT                          result in a deletion at this position"
FT   exon            44423..44691
FT                   /*tag=  aw
FT                   /number= 12
FT   intron          44692..47818
FT                   /*tag=  ax
FT                   /number= 12
FT   exon            47819..47897
FT                   /*tag=  ay
FT                   /number= 13
FT   intron          47898..50266
FT                   /*tag=  az
FT                   /number= 13
FT   variation       replace (48839,t)
FT                   /*tag=  ba
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            50267..50651
FT                   /*tag=  bb
FT                   /number= 14
FT   variation       replace (52265,g)
FT                   /*tag=  bc
FT                   /note= "SNP, single nucleotide polymorphism"
XX
PN   US2002142951-A1.
XX
PD   03-OCT-2002.
XX
PF   28-MAR-2001; 2001US-0818264.
XX
PR   28-MAR-2001; 2001US-0818264.
XX
PA   (WEBS/) WEBSTER M.
PA   (BEAS/) BEASLEY E M.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
XX
PI   Webster M, Beasley EM, Ketchum KA, Di Francesco V;
XX

  Query Match          75.0%;  Score 15;  DB 25;  Length 53226;
  Best Local Similarity  100.0%;  Pred. No. 27;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAG 16
              |||||||||||||||
Db      35513 CTTAGGAGGAACAAG 35527


RESULT 25
AAQ76831
ID   AAQ76831 standard; DNA; 125 BP.
XX
AC   AAQ76831;
XX
DT   25-MAR-2003  (updated)
DT   23-SEP-1994  (first entry)
XX
DE   Human genome fragment.
XX
KW   Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW   detection; homology; human; adrenal tissue; ds.
XX
OS   Homo sapiens.
XX
PN   WO9401548-A2.
XX
PD   20-JAN-1994.
XX
PF   13-JUL-1993;   93WO-GB01467.
XX
PR   13-JUL-1992;   92GB-0014857.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Gross J,  Hadfield KM,  Howells D,  Kelly M,  Shaw D;
PI   Sibson DR,  Starkey M;
XX
DR   WPI; 1994-035056/04.
XX
PT   New nucleic acid fragment encoding gene products - can be used
PT   for genetic analysis and mapping
XX
PS   Claim 1; Page 228; 616pp; English.
```

XX
CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (AAQ76401-Q77613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 125 BP; 29 A; 31 C; 39 G; 25 T; 1 other;

Query Match 70.0%; Score 14; DB 15; Length 125;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTC 18
||||||||||||||
Db 106 AGGAGGAACAAGTC 119

RESULT 26
ABX30187
ID ABX30187 standard; cDNA; 273 BP.
XX
AC ABX30187;
XX
DT 11-FEB-2003 (first entry)
XX
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #12244.
XX
KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.
XX
OS Homo sapiens.
XX
PN US2002110548-A1.
XX
PD 15-AUG-2002.
XX
PF 11-JUN-2001; 2001US-0878574.
XX
PR 22-NOV-1996; 96US-0753233.
PR 03-DEC-1997; 97US-0984246.
PR 09-SEP-1998; 98US-0149674.
PR 14-JUN-1999; 99US-0333177.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Sullivan F, Kriz R, Kumar R;
XX
DR WPI; 2003-066673/06.
XX
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
PT rejection -
XX
PS Disclosure; SEQ ID NO 12246; 6pp; English.
XX
CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
CC human GM4,6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 273 BP; 81 A; 57 C; 58 G; 77 T; 0 other;

Query Match 70.0%; Score 14; DB 25; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAA 15
||||||||||||||
Db 46 CTTAGGAGGAACAA 59

RESULT 27
AAX98399/c
ID AAX98399 standard; cDNA; 300 BP.
XX
AC AAX98399;
XX
DT 24-SEP-1999 (first entry)
XX
DE Human cancer cell derived cDNA #125.
XX
KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW integral membrane protein; aspartyl protease; GATA family; wnt family;
KW transcription factor; G-protein alpha subunit; protein phosphatase;
KW phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW protein kinase; tyrosine phosphatase; developmental signalling protein;
KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW prostate; ss.
XX
OS Homo sapiens.
XX
PN WO9933982-A2.
XX
PD 08-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US27610.
XX
PR 21-DEC-1998; 98US-0217471.
PR 23-DEC-1997; 97US-0068755.
PR 03-APR-1998; 98US-0080664.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-430243/36.
XX
PT New isolated human polynucleotides
XX
PS Claim 1; Page 348; 591pp; English.
XX
CC This invention describes novel isolated human polynucleotides obtained
CC by screening for differential expression in colon cancer, breast cancer
CC and lung cancer cell lines. The polynucleotides of the invention are

```
CC   represented in AAX98275-X99118 and encode polypeptides of protein
CC   families selected from 4 transmembrane segments integral membrane
CC   proteins, 7 transmembrane receptors, ATPases associated with various
CC   cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC   transcription factors, G-protein alpha subunit, phorbolesters or
CC   diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC   protein tyrosine phosphatase, trypsin, wnt family of developmental
CC   signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC   encoded polypeptides also have a functional domain selected from Ank
CC   repeat, basic region plus leucine zipper transcription factors,
CC   bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC   (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
CC   domain. The polynucleotides encode polypeptides with similarity to known
CC   protein families and are predicted to have similar properties. The novel
CC   polynucleotides can be used to develop products for use as therapeutic
CC   agents and in forensics, genetic analysis, mapping and diagnostic
CC   applications. In particular, the product can be used for the detection
CC   and management of cancers. They can be used for treating e.g. cervical
CC   cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC   retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC   myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC   myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC   hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC   epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC   mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC   prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC   the skin.
XX
SQ   Sequence 300 BP; 81 A; 35 C; 63 G; 121 T; 0 other;

  Query Match             70.0%;  Score 14;  DB 20;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACA 14
              ||||||||||||||
Db        230 CCTTAGGAGGAACA 217


RESULT 28
ABV06501
ID   ABV06501 standard; cDNA; 344 BP.
XX
AC   ABV06501;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 6492.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 1063; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 344 BP; 113 A; 52 C; 82 G; 94 T; 3 other;

  Query Match             70.0%;  Score 14;  DB 23;  Length 344;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTC 18
              ||||||||||||||
Db        100 AGGAGGAACAAGTC 113


RESULT 29
AAC03189
ID   AAC03189 standard; cDNA; 357 BP.
XX
AC   AAC03189;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 3187.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG03183.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 3187; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
```

```
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 357 BP; 112 A; 74 C; 78 G; 92 T; 1 other;

  Query Match             70.0%;  Score 14;  DB 21;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAGT 17
              ||||||||||||||
Db        223 TAGGAGGAACAAGT 236


RESULT 30
AAS38136/c
ID   AAS38136 standard; cDNA; 408 BP.
XX
AC   AAS38136;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #1194.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A,  Lamson G;
PI   Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones WL,  Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 896; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 408 BP; 57 A; 110 C; 140 G; 101 T; 0 other;

  Query Match             70.0%;  Score 14;  DB 22;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 GAGGAACAAGTCCC 20
              ||||||||||||||
Db         91 GAGGAACAAGTCCC 78


RESULT 31
AAX55494/c
ID   AAX55494 standard; DNA; 409 BP.
XX
AC   AAX55494;
XX
DT   27-JUL-1999  (first entry)
XX
DE   Rice Crm1 homologue DNA.
XX
KW   Pad1; Crm1; Jab1; AP-1 transcription factor activity; regulator; plant;
KW   maize; soybean; wheat; rice; yeast; human; isolation; transgenic; ss.
XX
OS   Oryza sativa.
XX
PN   WO9924574-A2.
XX
PD   20-MAY-1999.
XX
PF   04-NOV-1998;   98WO-US23487.
XX
PR   07-NOV-1997;   97US-0064914.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Anderson SL,  Hitz WD,  Kinney AJ,  Miao G;
PI   Morgante M,  Odell JT,  Sakai H;
XX
DR   WPI; 1999-327405/27.
DR   P-PSDB; AAY08447.
XX
PT   Plant homologues of yeast Pad1, Crm1 and human Jab1 and related
PT   polynucleotides
XX
PS   Claim 7; Page 42; 57pp; English.
XX
CC   This invention describes novel plant Pad1, Crm1 or Jab1 proteins which
CC   are capable of AP-1 transcription factor regulation. The proteins are
CC   thought to interact with transcription factors altering gene expression.
CC   The nucleic acid sequences of the invention may be used to isolate cDNAs
CC   and genes encoding homologous proteins from the same or other plant
CC   species. Synthetic peptides of the proteins may be synthesized to
CC   generate antibodies that are useful for screening expression libraries.
CC   Transgenic plants may be produced using the nucleic acid sequences to
CC   alter the levels of Pad1, Crm1 and Jab1 present in the plants. Altering
CC   the levels of these proteins would alter the level of AP-1 transcription
CC   factor activity in the plants.
XX
SQ   Sequence 409 BP; 119 A; 90 C; 82 G; 118 T; 0 other;

  Query Match             70.0%;  Score 14;  DB 20;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAA 15
              ||||||||||||||
Db        296 CTTAGGAGGAACAA 283


RESULT 32
AAS26323
ID   AAS26323 standard; cDNA; 442 BP.
XX
AC   AAS26323;
XX
DT   07-NOV-2001  (first entry)
```

XX
DE Human cDNA encoding a novel secreted protein, Seq ID 502.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR P-PSDB; AAU16336.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; SEQ ID No 502; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 70.0%; Score 14; DB 22; Length 442;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TAGGAGGAACAAGT 17
||||||||||||||
Db 141 TAGGAGGAACAAGT 154

RESULT 33
ABX73664
ID ABX73664 standard; DNA; 442 BP.
XX
AC ABX73664;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polynucleotide #492.
XX
KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-0764864.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR P-PSDB; ABU55404.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for

```
PT   treating, inhibiting or preventing e.g. neural, immune system,
PT   muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT   cardiovascular or renal disorders -
XX
PS   Claim 1; SEQ ID NO 502; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC   arthritis and multiple sclerosis), muscular disorders, respiratory
CC   diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC   reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC   cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC   anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC   kidney failure and end-stage renal disease), hyperproliferative disorders
CC   (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC   septic shock, bursitis and appendicitis), allergic reactions and
CC   conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC   atherosclerosis and myocardial infarction) and cancerous diseases.
CC   Sequences ABX73173-ABX74167 represent human novel polynucleotides of the
CC   invention.
XX
SQ   Sequence 442 BP; 192 A; 70 C; 83 G; 94 T; 3 other;

  Query Match             70.0%;  Score 14;  DB 25;  Length 442;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAGT 17
              ||||||||||||||
Db        141 TAGGAGGAACAAGT 154


RESULT 34
ABV36454
ID   ABV36454 standard; cDNA; 467 BP.
XX
AC   ABV36454;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 36445.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 7534; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 467 BP; 148 A; 84 C; 117 G; 118 T; 0 other;

  Query Match             70.0%;  Score 14;  DB 23;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTC 18
              ||||||||||||||
Db        181 AGGAGGAACAAGTC 194


RESULT 35
ABA43976/c
ID   ABA43976 standard; DNA; 477 BP.
XX
AC   ABA43976;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human breast cell single exon nucleic acid probe #2671.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes,
PT   useful for measuring gene expression in sample derived from human
PT   breast, comprises number of single exon nucleic acid probes  -
XX
PS   Claim 1; SEQ ID NO 2671; 327pp + sequence listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting
CC   the probes with a collection of detectably labelled nucleic acids
```

CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 477 BP; 112 A; 93 C; 127 G; 145 T; 0 other;

Query Match 70.0%; Score 14; DB 22; Length 477;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GAGGAACAAGTCCC 20
||||||||||||||
Db 314 GAGGAACAAGTCCC 301

Search completed: August 14, 2003, 21:41:21
Job time : 130 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:41:37 ; Search time 1126.8 Seconds
(without alignments)
388.250 Million cell updates/sec

Title: US-10-074-620-1
Perfect score: 18
Sequence: 1 ggctggtgtcacctgtta 18

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		17	94.4	927	10	BF536318	BF536318 602051968
2		17	94.4	1080	29	CNS02AHD	AL188554 Tetraodon
c 3		16	88.9	124	12	BI014294	BI014294 MR4-ET014
c 4		16	88.9	170	9	AA294899	AA294899 EST100142
c 5		16	88.9	183	10	BF170207	BF170207 PCL0425 M
c 6		16	88.9	199	9	AW405433	AW405433 UI-HF-BL0
c 7		16	88.9	221	10	BF354113	BF354113 PM4-HT072
c 8		16	88.9	273	10	BE969911	BE969911 601679494
c 9		16	88.9	295	9	AA339056	AA339056 EST44112
c 10		16	88.9	300	9	AU099544	AU099544 AU099544
11		16	88.9	301	10	BF516504	BF516504 UI-H-BW1-
c 12		16	88.9	317	14	CB267362	CB267362 1006268 H
c 13		16	88.9	319	13	BQ369040	BQ369040 PM3-GN051
c 14		16	88.9	319	13	BQ369218	BQ369218 PM3-GN051
c 15		16	88.9	335	10	BF850266	BF850266 CM3-EN007
c 16		16	88.9	377	14	CB130696	CB130696 K-EST0180
c 17		16	88.9	394	14	H69706	H69706 yr93a08.s1
c 18		16	88.9	395	12	BM789452	BM789452 K-EST0069
c 19		16	88.9	404	10	BF668175	BF668175 602122957
c 20		16	88.9	413	14	CB270357	CB270357 1009264 H
c 21		16	88.9	428	9	AA135078	AA135078 zo26d06.r
c 22		16	88.9	430	9	AW732816	AW732816 bb14h10.y
c 23		16	88.9	431	10	BF850584	BF850584 PM1-EN006
c 24		16	88.9	431	14	CB270361	CB270361 1009268 H
c 25		16	88.9	433	14	CB145604	CB145604 K-EST0200
c 26		16	88.9	434	14	CB129626	CB129626 K-EST0179
c 27		16	88.9	436	14	CB142213	CB142213 K-EST0195
c 28		16	88.9	438	10	BF858511	BF858511 RC1-FT019
c 29		16	88.9	447	10	BG104409	BG104409 602311027
c 30		16	88.9	449	10	BF130453	BF130453 601818761
c 31		16	88.9	452	10	BE909128	BE909128 601501730
c 32		16	88.9	452	13	BX475327	BX475327 DKFZp686I
c 33		16	88.9	453	9	AA496839	AA496839 ae33b02.r
34		16	88.9	454	12	BM171994	BM171994 imageqc_3
c 35		16	88.9	455	10	BF035036	BF035036 601456130
c 36		16	88.9	459	10	BF857906	BF857906 RC1-FT019
c 37		16	88.9	466	14	CB128372	CB128372 K-EST0177
c 38		16	88.9	473	10	BF669564	BF669564 602120279
c 39		16	88.9	473	13	BX095645	BX095645 BX095645
c 40		16	88.9	481	9	AW250575	AW250575 2821885.5
c 41		16	88.9	484	9	AW246279	AW246279 2821886.5
c 42		16	88.9	484	10	BE539568	BE539568 601060282
c 43		16	88.9	487	14	CB140784	CB140784 K-EST0194
c 44		16	88.9	490	10	BF922602	BF922602 QV4-NT025
c 45		16	88.9	490	13	BX471656	BX471656 DKFZp686L
c 46		16	88.9	494	10	BF857889	BF857889 RC1-FT019
c 47		16	88.9	501	12	BM795352	BM795352 K-EST0077
c 48		16	88.9	501	13	BX280514	BX280514 BX280514
c 49		16	88.9	509	10	BF971748	BF971748 602240020
c 50		16	88.9	511	10	BF665487	BF665487 602123870
c 51		16	88.9	512	10	BF699171	BF699171 602126810
c 52		16	88.9	512	10	BE958443	BE958443 601644752
c 53		16	88.9	520	9	AW328258	AW328258 ds01c08.x
c 54		16	88.9	525	10	BE390761	BE390761 601287253
c 55		16	88.9	533	10	BE799191	BE799191 601592752
c 56		16	88.9	540	9	AA307552	AA307552 EST178632
c 57		16	88.9	542	14	CB144397	CB144397 K-EST0198
c 58		16	88.9	542	14	CB269294	CB269294 1008201 H
c 59		16	88.9	549	10	BG495444	BG495444 602539755
c 60		16	88.9	550	14	CB128792	CB128792 K-EST0178
c 61		16	88.9	551	13	BX475673	BX475673 DKFZp686H
c 62		16	88.9	551	14	CB266268	CB266268 1005173 H
c 63		16	88.9	557	10	BE389991	BE389991 601285456
c 64		16	88.9	561	10	BE266170	BE266170 601191319
c 65		16	88.9	570	10	BG687435	BG687435 602639402
66		16	88.9	572	9	AA203223	AA203223 zx56e02.r
c 67		16	88.9	581	10	BG528181	BG528181 602557578
c 68		16	88.9	581	12	BI915162	BI915162 603177375
c 69		16	88.9	584	10	BE279400	BE279400 601157972
c 70		16	88.9	586	9	AW675194	AW675194 bb41a06.y
c 71		16	88.9	588	10	BF701069	BF701069 602128125
c 72		16	88.9	592	12	BM048988	BM048988 603621859
c 73		16	88.9	596	10	BE733288	BE733288 601569464
c 74		16	88.9	602	10	BE298939	BE298939 601119850
c 75		16	88.9	603	10	BE266714	BE266714 601190278
76		16	88.9	606	10	BG529561	BG529561 602557970
c 77		16	88.9	608	9	AU123424	AU123424 AU123424

c 78	16	88.9	608	12	BI258411	BI258411 602972681
c 79	16	88.9	617	12	BI856915	BI856915 603384737
c 80	16	88.9	618	10	BE251364	BE251364 601116217
c 81	16	88.9	626	12	BG777445	BG777445 602664752
c 82	16	88.9	626	13	BX475535	BX475535 DKFZp686E
c 83	16	88.9	627	10	BF665573	BF665573 602119773
c 84	16	88.9	629	10	BG529643	BG529643 602558076
c 85	16	88.9	632	10	BB664539	BB664539 BB664539
c 86	16	88.9	633	10	BG746738	BG746738 602704066
c 87	16	88.9	639	10	BE741309	BE741309 601594146
c 88	16	88.9	640	10	BF665524	BF665524 602119718
c 89	16	88.9	640	12	BI521439	BI521439 603081239
c 90	16	88.9	641	12	BI334698	BI334698 602998678
c 91	16	88.9	643	10	BG506929	BG506929 601861554
c 92	16	88.9	643	10	BB614497	BB614497 BB614497
c 93	16	88.9	647	10	BF691465	BF691465 602247634
c 94	16	88.9	650	10	BE799793	BE799793 601591396
c 95	16	88.9	651	10	BE615490	BE615490 601280992
c 96	16	88.9	651	10	BE958171	BE958171 601644629
c 97	16	88.9	652	14	Z99383	Z99383 HSZ99383 DK
c 98	16	88.9	653	10	BF106351	BF106351 601823571
c 99	16	88.9	655	10	BE548866	BE548866 601071888
c 100	16	88.9	656	10	BG284498	BG284498 602408640
c 101	16	88.9	657	10	BE394893	BE394893 601310655
c 102	16	88.9	659	10	BF695670	BF695670 601852595
c 103	16	88.9	661	12	BG769924	BG769924 602744796
c 104	16	88.9	661	12	BI834955	BI834955 603088836
c 105	16	88.9	667	10	BF316186	BF316186 601896192
c 106	16	88.9	676	12	BG898948	BG898948 HOA4-1-E-
c 107	16	88.9	678	10	BE390209	BE390209 601285755
c 108	16	88.9	678	10	BE618501	BE618501 601462941
c 109	16	88.9	678	12	BG777022	BG777022 602664227
c 110	16	88.9	679	13	BQ109463	BQ109463 imageqc_7
c 111	16	88.9	680	10	BG504177	BG504177 602552709
c 112	16	88.9	683	10	BE728332	BE728332 601561721
c 113	16	88.9	684	10	BE387969	BE387969 601282740
c 114	16	88.9	687	10	BG496927	BG496927 602541345
c 115	16	88.9	692	10	BF667533	BF667533 602120890
c 116	16	88.9	693	10	BG471279	BG471279 602512518
c 117	16	88.9	694	10	BF125491	BF125491 601763433
c 118	16	88.9	702	10	BF185178	BF185178 601843846
c 119	16	88.9	703	12	BI258980	BI258980 602971630
c 120	16	88.9	705	10	BE277837	BE277837 601120048

ALIGNMENTS

```
RESULT 1
BF536318
LOCUS       BF536318                927 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  602051968F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4191040 5',
            mRNA sequence.
ACCESSION   BF536318
VERSION     BF536318.1  GI:11623686
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9519  row: h  column: 17
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4191040"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_SG2"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      221 a    224 c    266 g    214 t      2 others
ORIGIN

  Query Match            94.4%;  Score 17;  DB 10;  Length 927;
  Best Local Similarity  100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db        818 GCTGGTGTCACCTGTTA 834


RESULT 2
CNS02AHD
LOCUS       CNS02AHD                1080 bp    DNA    linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            251P01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL188554
VERSION     AL188554.1  GI:7826658
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1080)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Tetraodon nigroviridis"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="251P01"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG251CH01LP1~end : T7"
BASE COUNT       257 a    267 c    223 g    324 t      9 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 29;  Length 1080;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db        324 GCTGGTGTCACCTGTTA 340


RESULT 3
BI014294/c
LOCUS       BI014294                124 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  MR4-ET0142-310101-001-h11 ET0142 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI014294
VERSION     BI014294.1  GI:14418365
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-ET0142-
            310101-001-h11&t3=2001-01-31&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 124.
FEATURES             Location/Qualifiers
     source          1. .124
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ET0142"
                     /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       26 a     51 c     19 g     28 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 12;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         90 GGCTGGTGTCACCTGT 75


RESULT 4
AA294899/c
LOCUS       AA294899                170 bp    mRNA    linear   EST 18-APR-1997
DEFINITION  EST100142 Pancreas tumor I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA294899
VERSION     AA294899.1  GI:1947334
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
   PUBMED   7566098
COMMENT     Other_ESTs: THC172745
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .170
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):190475"
                     /db_xref="taxon:9606"
                     /dev_stage="adult"
                     /clone_lib="Pancreas tumor I"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       39 a     48 c     51 g     28 t      4 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 51;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
```

```
Db        135 GGCTGGTGTCACCTGT 120


RESULT 5
BF170207/c
LOCUS       BF170207                183 bp    mRNA    linear   EST 23-MAR-2001
DEFINITION  PCL0425 Myeloma (PCL) cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BF170207
VERSION     BF170207.1  GI:13436309
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183)
  AUTHORS   Claudio,J.O., Masih-Khan,E., Tang,H., Goncalves,J., Voralia,M., Li
            ,Z.H., Nadeem,V., Cukerman,E., Franciso-Pabalan,O., Liew,C.C.,
            Woodgett,J.R. and Stewart,A.K.
  TITLE     A molecular compendium of genes expressed in multiple myeloma
  JOURNAL   Blood 100 (6), 2175-2186 (2002)
  MEDLINE   22188429
   PUBMED   12200383
COMMENT     Contact: A. Keith Stewart, M.D.
            Oncology Research
            University Health Network
            610 University Ave., 5-126, Toronto, Ontario, M5G 2M9, Canada
            Tel: (416) 946-4639
            Fax: (416) 946-6546
            Email: k.stewart@utoronto.ca
            PCR PRimers
            FORWARD: 5'-GCCAAGCTCGAAATTAACCCTCACTAAAGGG-3'
            BACKWARD: 5'-CCAGTGAATTGTAATACGACTCACTATAGGGCG-3'
            Seq primer: 5'-GAAATTAACCCTCACTAAAGG-3'.
FEATURES             Location/Qualifiers
     source          1. .183
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /tissue_type="Blood"
                     /cell_type="myeloma"
                     /dev_stage="Plasma cell leukemia"
                     /clone_lib="Myeloma (PCL) cDNA library"
                     /note="Vector: Lambda Zap Express; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was purified from plasma cell leukemia
                     patient's peripheral blood containing >95% myeloma. An
                     oligo d(T)18 primer containing XhoI restriction site was
                     used to prime first strand synthesis using M-MLV reverse
                     transcriptase. To protect the cDNAs from XhoI digestion in
                     subsequent cloning step, the nucleotide analogue
                     5-methyl-dCTP was added to the nucleotide mixture and
                     [a-32P]dATP was added to monitor the quantity and quality
                     of first strand synthesis. After second-strand synthesis
                     and blunting of cDNA termini, EcoRI adapters were ligated
                     , followed by kinase treatment and digestion with XhoI.
                     The cDNAs were then size-fractionated using Sephacryl
                     S-500 column and then ligated into EcoRI and XhoI digested
                     Lambda Zap Express vector. The ligation product was
                     packaged using Gigapack II packaging extract. The library
                     had primary titre of approx. 1x106. Clones from the
                     primary library were randomly selected for single pass
                     sequencing."
BASE COUNT       41 a     49 c     56 g     37 t
ORIGIN

  Query Match           88.9%;  Score 16;  DB 10;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        143 GGCTGGTGTCACCTGT 128


RESULT 6
AW405433/c
LOCUS       AW405433                199 bp    mRNA    linear   EST 16-FEB-2000
DEFINITION  UI-HF-BL0-adc-h-06-0-UI.r1 NIH_MGC_37 Homo sapiens cDNA clone
            IMAGE:3061115 5', mRNA sequence.
ACCESSION   AW405433
VERSION     AW405433.1  GI:6924490
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .199
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3061115"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /clone_lib="NIH_MGC_37"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (1.5-2.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT       60 a     57 c     53 g     29 t
ORIGIN

  Query Match           88.9%;  Score 16;  DB 9;  Length 199;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         70 GGCTGGTGTCACCTGT 55


RESULT 7
BF354113/c
LOCUS       BF354113                221 bp    mRNA    linear   EST 22-NOV-2000
DEFINITION  PM4-HT0726-270500-002-c07 HT0726 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF354113
VERSION     BF354113.1  GI:11313187
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 221)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
```

```
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
  TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
  JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE 20202663
   PUBMED 10737800
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-HT0726-
          270500-002-c07&t3=2000-05-27&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 28
          High quality sequence stop: 219.
FEATURES             Location/Qualifiers
     source          1. .221
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0726"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       39 a     94 c     31 g     57 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 221;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         99 GGCTGGTGTCACCTGT 84


RESULT 8
BE969911/c
LOCUS       BE969911                273 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  601679494F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949966 5',
          mRNA sequence.
ACCESSION   BE969911
VERSION     BE969911.1  GI:10582844
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: CLONETECH Laboratories, Inc.
           cDNA Library Preparation: CLONETECH Laboratories, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM815  row: k  column: 23
          High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .273
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949966"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_78"
                     /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.2 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT       74 a     68 c     85 g     46 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        151 GGCTGGTGTCACCTGT 136


RESULT 9
AA339056/c
LOCUS       AA339056                295 bp    mRNA    linear   EST 21-APR-1997
DEFINITION  EST44112 Fetal brain I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA339056
VERSION     AA339056.1  GI:1991304
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
          ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
          ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
          Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
          ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
          Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
          Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
          Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
          Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
          Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
          Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
          Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
          , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
          Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
          Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
          ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
          Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
          based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
   PUBMED   7566098
COMMENT   Other_ESTs: THC172745
          Contact: Kerlavage, AR
          Bioinformatics
          The Institute for Genomic Research
          9712 Medical Center Drive, Rockville, MD 20850 USA
          Tel: 3018699056
          Fax: 3018699423
```

```
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):140673"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /dev_stage="fetus, 24 wks"
                     /clone_lib="Fetal brain I"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       77 a     74 c     89 g     53 t      2 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 295;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        161 GGCTGGTGTCACCTGT 146


RESULT 10
AU099544/c
LOCUS       AU099544                300 bp    mRNA    linear   EST 05-APR-2001
DEFINITION  AU099544 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HSI08640 similar to Human translation initiation factor eIF3 p44
            subunit mRNA, mRNA sequence.
ACCESSION   AU099544
VERSION     AU099544.1  GI:13550673
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
  TITLE     In silico mapping of the 5'-ends of human mRNAs using full-length
            enriched and 5'-end enriched cDNA libraries constructed by
            Oligo-capping method
  JOURNAL   Unpublished
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HSI08640"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       74 a     77 c     95 g     54 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        186 GGCTGGTGTCACCTGT 171


RESULT 11
BF516504
LOCUS       BF516504                301 bp    mRNA    linear   EST 07-DEC-2000
DEFINITION  UI-H-BW1-aoc-g-07-0-UI.s1 NCI_CGAP_Sub7 Homo sapiens cDNA clone
            IMAGE:3084469 3', mRNA sequence.
ACCESSION   BF516504
VERSION     BF516504.1  GI:11601683
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NCI-CGAP clone distribution
            information can be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3084469"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Sub7"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub7
                     is a subtracted library derived from NCI_CGAP_Sub6.  The
                     NCI_CGAP_Sub7 library had 12 million recombinants.  A
                     single-stranded DNA preparation of NCI_CGAP_Sub6 was used
                     as a tracer in a subtractive  hybridization with a driver
                     comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1  LLAM
                     3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
                     1322376-1323911,  1456008-1456775,1500552-1502855);
                     NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778
                     (IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
                     1492104-1493255);  NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
                     3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-1522439
                     ); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
                     3733-3735  (IMAGE CloneIDs 1257096-1258631,1469064-1470983
                     , 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
                     2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
                     1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
                     LLAM 2644-2653,  2871-2872 (IMAGE CloneIDs 1057416-1061255
                     ,1144584-1145351). (6% of the driver  population), plus a
                     pool of 3,840 arrayed clones from  NCI_CGAP_Sub1 (IMAGE
                     CloneIDs 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
                     CloneIDs 2710536-2712455)    (4% of the driver population
                     ), plus a pool of 11,136 clones from NCI_CGAP_Sub3  (IMAGE
                     CloneIDs 2712456-2723591) (10% of the driver population),
                     plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
                     CloneIDs 2723592-2729326) (40% of the driver  population),
                     plus a pool of 4032 clones from NCI_CGAP_Sub6 (IMAGE
                     CloneIDs 2728969-2733190) (40% of the driver population).
                     Subtraction was performed as previously described [Bonaldo
                     , Lennon & Soares (1996): Normalization and  Subtraction:
                     Two Approaches To Facilitate Gene Discovery. Genome
                     Research 6, 791-806.
                     TAG_SEQ=None found"
BASE COUNT       44 a     88 c     82 g     87 t
ORIGIN
```

```
  Query Match             88.9%;  Score 16;  DB 10;  Length 301;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        229 GGCTGGTGTCACCTGT 244


RESULT 12
CB267362/c
LOCUS       CB267362                317 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  1006268 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CB267362
VERSION     CB267362.1  GI:28441947
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT       89 a     78 c     97 g     52 t      1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        136 GGCTGGTGTCACCTGT 121


RESULT 13
BQ369040/c
LOCUS       BQ369040                319 bp    mRNA    linear   EST 21-MAY-2002
DEFINITION  PM3-GN0516-060601-031-g03 GN0516 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ369040
VERSION     BQ369040.1  GI:21044554
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-GN0516-
            060601-031-g03&t3=2001-06-06&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="GN0516"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT       93 a     79 c     92 g     54 t      1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 13;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         97 GGCTGGTGTCACCTGT 82


RESULT 14
BQ369218/c
LOCUS       BQ369218                319 bp    mRNA    linear   EST 21-MAY-2002
DEFINITION  PM3-GN0516-020701-031-g03 GN0516 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ369218
VERSION     BQ369218.1  GI:21044732
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
```

```
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-GN0516-
            020701-031-g03&t3=2001-07-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 319.
FEATURES              Location/Qualifiers
     source           1. .319
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="GN0516"
                      /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                      ; Site_2: SmaI; A mini-library was made by cloning
                      products derived from  ORESTES PCR (U.S. Letters Patent
                      application No. 196,716 -  Ludwig Institute for Cancer
                      Research) profiles into  the pUC 18 vector. Reverse
                      transcription of tissue mRNA and cDNA amplification were
                      performed under low stringency conditions."
BASE COUNT       93 a      79 c      92 g      54 t       1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 13;  Length 319;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         97 GGCTGGTGTCACCTGT 82


RESULT 15
BF850266/c
LOCUS       BF850266                335 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  CM3-EN0077-181100-489-d09 EN0077 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF850266
VERSION     BF850266.1  GI:12237428
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-EN0077-
            181100-489-d09&t3=2000-11-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 52
            High quality sequence stop: 335.
FEATURES              Location/Qualifiers
     source           1. .335
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="EN0077"
                      /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       69 a      96 c     107 g      63 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 335;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        335 GGCTGGTGTCACCTGT 320


RESULT 16
CB130696/c
LOCUS       CB130696                377 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0180649 L11SNU354 Homo sapiens cDNA clone L11SNU354-17-C06 5',
            mRNA sequence.
ACCESSION   CB130696
VERSION     CB130696.1  GI:28095037
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 17  row: C  column: 06
            High quality sequence stop: 377.
FEATURES              Location/Qualifiers
     source           1. .377
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="L11SNU354-17-C06"
                      /sex="M"
                      /tissue_type="Liver"
                      /cell_type="Polygonal"
                      /cell_line="SNU-354"
                      /lab_host="Top10F'"
                      /clone_lib="L11SNU354"
                      /note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
                      Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                      bacterial alkaline phosphatase (BAP) and then decapped
```

with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

```
BASE COUNT      104 a    102 c    107 g     64 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 377;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        141 GGCTGGTGTCACCTGT 126
```

RESULT 17
H69706/c

```
LOCUS       H69706                  394 bp    mRNA    linear   EST 24-OCT-1995
DEFINITION  yr93a08.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:212822 3', mRNA sequence.
ACCESSION   H69706
VERSION     H69706.1  GI:1039912
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 821
            High quality sequence stops: 345
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 821   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3785663"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:212822"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      102 a     81 c    106 g    102 t      3 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 394;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 CTGGTGTCACCTGTTA 18
              ||||||||||||||||
Db        239 CTGGTGTCACCTGTTA 224
```

RESULT 18
BM789452/c

```
LOCUS       BM789452                395 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0069040 S22SNU16 Homo sapiens cDNA clone S22SNU16-2-A08 5',
            mRNA sequence.
ACCESSION   BM789452
VERSION     BM789452.1  GI:19137684
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 2  row: A  column: 08
            High quality sequence stop: 395.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S22SNU16-2-A08"
                     /sex="F"
                     /tissue_type="Ascites"
                     /cell_type="Lymphoblast-like"
                     /cell_line="SNU-16"
                     /lab_host="DH10B"
                     /clone_lib="S22SNU16"
                     /note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The S22SNU16 library was contributed by the
                     Soares laboratory and it was constructed as described by
                     Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
                     Research 6(9): 791-806. RNA was prepared from harvested
                     cells of SNU-16 culture. SNU-16 cell was obtained from
                     Korean Cell Line Bank (KCLB). SNU-16 was established from
                     ascitic fluids of Korean patients by Park J.G. et al.
                     (1990), Cancer Res 50: 2773-2780."
BASE COUNT      111 a    104 c    116 g     64 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 12;  Length 395;
```

```
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        111 GGCTGGTGTCACCTGT 96


RESULT 19
BF668175/c
LOCUS       BF668175                404 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  602122957F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4279958 5',
            mRNA sequence.
ACCESSION   BF668175
VERSION     BF668175.1  GI:11942070
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1106  row: a  column: 15
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4279958"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_56"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      108 a    103 c    125 g     67 t      1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 404;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        154 GGCTGGTGTCACCTGT 139


RESULT 20
CB270357/c
LOCUS       CB270357                413 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  1009264 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CB270357
VERSION     CB270357.1  GI:28444942
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT      114 a    104 c    119 g     71 t      5 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        131 GGCTGGTGTCACCTGT 116


RESULT 21
AA135078/c
LOCUS       AA135078                428 bp    mRNA    linear   EST 06-AUG-1997
DEFINITION  zo26d06.r1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:588011 5', mRNA sequence.
ACCESSION   AA135078
VERSION     AA135078.1  GI:1696180
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```

```
            Insert Length: 1117   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 263.
FEATURES              Location/Qualifiers
     source           1. .428
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:4620313"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:588011"
                      /tissue_type="tumor"
                      /cell_line="T84 carcinoma cell line"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /clone_lib="Stratagene colon (#937204)"
                      /note="Organ: colon; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. T-84 colonic epithelial cell line.  Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      128 a    112 c    113 g     71 t      4 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 428;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         67 GGCTGGTGTCACCTGT 52


RESULT 22
AW732816/c
LOCUS       AW732816                 430 bp    mRNA    linear   EST 21-APR-2000
DEFINITION  bb14h10.y1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2962915 5'
            similar to TR:O14801 O14801 EUKARYOTIC TRANSLATION INITIATION
            FACTOR 3 SUBUNIT. [1] ;, mRNA sequence.
ACCESSION   AW732816
VERSION     AW732816.1  GI:7633154
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40RP from Gibco
            High quality sequence stop: 412.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2962915"
                      /tissue_type="choriocarcinoma"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_21"
                      /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                      Site_2: EcoRI; cDNA made by oligo-dT priming.
                      Directionally cloned into EcoRI/XhoI sites using the
                      following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                      for average insert size 1.8kb.  Library constructed by
                      Ling Hong in the laboratory of Gerald M. Rubin (University
                      of California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      119 a    112 c    125 g     74 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        140 GGCTGGTGTCACCTGT 125


RESULT 23
BF850584/c
LOCUS       BF850584                 431 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  PM1-EN0065-191100-007-h05 EN0065 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF850584
VERSION     BF850584.1  GI:12237746
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM1&t2=PM1-EN0065-
            191100-007-h05&t3=2000-11-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 39
            High quality sequence stop: 431.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="EN0065"
                      /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       65 a    157 c    107 g    102 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 66;
```

```
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        361 GGCTGGTGTCACCTGT 346


RESULT 24
CB270361/c
LOCUS       CB270361                431 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  1009268 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CB270361
VERSION     CB270361.1  GI:28444946
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT      120 a    112 c    125 g     74 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        135 GGCTGGTGTCACCTGT 120


RESULT 25
CB145604/c
LOCUS       CB145604                433 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0200441 L11SNU354s1 Homo sapiens cDNA clone L11SNU354s1-13-C11
            5', mRNA sequence.
ACCESSION   CB145604
VERSION     CB145604.1  GI:28125676
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 13  row: C  column: 11
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .433
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L11SNU354s1-13-C11"
                     /sex="M"
                     /tissue_type="Liver"
                     /cell_type="Polygonal"
                     /cell_line="SNU-354"
                     /lab_host="Top10F'"
                     /clone_lib="L11SNU354s1"
                     /note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including
                     EcoRI site by treatment of T4 RNA ligase and the first
                     strand cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama-Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library. After analyzing and
                     sequencing about 2,000 - 3,000 colonies in original cDNA
                     library, the abundant cDNAs were selected and amplified by
                     PCR reaction using vector region primer including T7
                     promotor as 5' primer and N(dT)14 as 3' primer. The PCR
                     products were used as template for synthesis of
                     biotinylated single stranded RNA by in vitro transcription
                     reaction. The synthesized RNA probes were hybridized with
                     antisense single stranded cDNAs prepared from original
                     liberary and incubated with avidin-gel. After removing
                     DNA-RNA hybrids by centrifuge, the subtracted cDNA
                     libraries were constructed by transformaion of the
                     remaining DNA into competent cells E. coli Top10F' with
                     electroporation method."
BASE COUNT      110 a    115 c    131 g     77 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 433;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        187 GGCTGGTGTCACCTGT 172


RESULT 26
CB129626/c
LOCUS       CB129626                434 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0179303 L11SNU354 Homo sapiens cDNA clone L11SNU354-16-B02 5',
            mRNA sequence.
ACCESSION   CB129626
VERSION     CB129626.1  GI:28093354
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 16  row: B  column: 02
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L11SNU354-16-B02"
                     /sex="M"
                     /tissue_type="Liver"
                     /cell_type="Polygonal"
                     /cell_line="SNU-354"
                     /lab_host="Top10F'"
                     /clone_lib="L11SNU354"
                     /note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including
                     EcoRI site by treatment of T4 RNA ligase and the first
                     strand cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama-Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT      122 a    112 c    125 g     75 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        142 GGCTGGTGTCACCTGT 127


RESULT 27
CB142213/c
LOCUS       CB142213                436 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0195953 L11SNU354s1 Homo sapiens cDNA clone L11SNU354s1-19-E11
            5', mRNA sequence.
ACCESSION   CB142213
VERSION     CB142213.1  GI:28118094
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 19  row: E  column: 11
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .436
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L11SNU354s1-19-E11"
                     /sex="M"
                     /tissue_type="Liver"
                     /cell_type="Polygonal"
                     /cell_line="SNU-354"
                     /lab_host="Top10F'"
                     /clone_lib="L11SNU354s1"
                     /note="Organ: Liver; Vector: pCNS-D2; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including
                     EcoRI site by treatment of T4 RNA ligase and the first
                     strand cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama-Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library. After analyzing and
                     sequencing about 2,000 - 3,000 colonies in original cDNA
                     library, the abundant cDNAs were selected and amplified by
                     PCR reaction using vector region primer including T7
                     promotor as 5' primer and N(dT)14 as 3' primer. The PCR
                     products were used as template for synthesis of
                     biotinylated single stranded RNA by in vitro transcription
                     reaction. The synthesized RNA probes were hybridized with
                     antisense single stranded cDNAs prepared from original
                     liberary and incubated with avidin-gel. After removing
                     DNA-RNA hybrids by centrifuge, the subtracted cDNA
                     libraries were constructed by transformaion of the
                     remaining DNA into competent cells E. coli Top10F' with
                     electroporation method."
BASE COUNT      112 a    113 c    132 g     79 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 436;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        172 GGCTGGTGTCACCTGT 157


RESULT 28
BF858511/c
LOCUS       BF858511                438 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  RC1-FT0190-221100-021-c02 FT0190 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF858511
VERSION     BF858511.1  GI:12246255
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-FT0190-
            221100-021-c02&t3=2000-11-22&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 27
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="FT0190"
                     /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      115 a    110 c    133 g     80 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 438;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        173 GGCTGGTGTCACCTGT 158


RESULT 29
BG104409/c
LOCUS       BG104409                447 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602311027F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4422918 5',
            mRNA sequence.
ACCESSION   BG104409
VERSION     BG104409.1  GI:12598251
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1218  row: f  column: 07
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4422918"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      129 a    116 c    134 g     68 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 447;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        119 GGCTGGTGTCACCTGT 104


RESULT 30
BF130453/c
LOCUS       BF130453                449 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  601818761F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4044398 5',
            mRNA sequence.
ACCESSION   BF130453
VERSION     BF130453.1  GI:10969493
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM868  row: j  column: 15
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4044398"
                     /tissue_type="hypernephroma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_58"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
```

SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
BASE COUNT 110 a 124 c 130 g 85 t
ORIGIN

Query Match 88.9%; Score 16; DB 10; Length 449;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 122 GGCTGGTGTCACCTGT 107

RESULT 31
BE909128/c
LOCUS BE909128 452 bp mRNA linear EST 20-OCT-2000
DEFINITION 601501730F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3903510 5', mRNA sequence.
ACCESSION BE909128
VERSION BE909128.1 GI:10404401
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9707 row: l column: 07
High quality sequence stop: 452.
FEATURES Location/Qualifiers
source 1. .452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3903510"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."
BASE COUNT 127 a 117 c 133 g 75 t
ORIGIN

Query Match 88.9%; Score 16; DB 10; Length 452;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 137 GGCTGGTGTCACCTGT 122

RESULT 32
BX475327/c
LOCUS BX475327 452 bp mRNA linear EST 12-JUN-2003
DEFINITION DKFZp686I09176_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686I09176 5', mRNA sequence.
ACCESSION BX475327
VERSION BX475327.1 GI:31672610
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
JOURNAL Unpublished
COMMENT Contact: Bahr A
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKFZp686I09176) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers
source 1. .452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686I09176"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"
BASE COUNT 114 a 120 c 138 g 80 t
ORIGIN

Query Match 88.9%; Score 16; DB 13; Length 452;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 184 GGCTGGTGTCACCTGT 169

RESULT 33
AA496839/c
LOCUS AA496839 453 bp mRNA linear EST 12-AUG-1997
DEFINITION ae33b02.r1 Gessler Wilms tumor Homo sapiens cDNA clone IMAGE:897579 5' similar to contains Alu repetitive element;contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION AA496839
VERSION AA496839.1 GI:2230160
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997

```
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:897579"
                     /sex="pooled (6)"
                     /lab_host="DH10B"
                     /clone_lib="Gessler Wilms tumor"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; RNA
                     was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
                     RNA was prepared by acid-phenol, followed by one round of
                     oligo dT selection.  cDNA library preparation was with
                     the BRL/Life Tech. Superscript Plasmid system.  An
                     oligo-dT NotI primer for first strand synthesis generated
                     gcggccgccc(t)n at the 3' end of the clones.  A 5' SalI
                     adaptor was used  with sequence 5'-gtcgacccacgcgtccg-3'.
                     Resulting cDNAs were size selected (average size 2 kb),
                     NotI digested, and ligated into NotI/SalI-cut pSPORT1.
                     Library was constructed by Dr. Manfred Gessler."
BASE COUNT      115 a    124 c    135 g     79 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 453;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        117 GGCTGGTGTCACCTGT 102


RESULT 34
BM171994
LOCUS       BM171994                454 bp    mRNA    linear   EST 04-DEC-2001
DEFINITION  imageqc_3_2001/smm196bdff41.x1 NIH_MGC_61 Homo sapiens cDNA clone
            IMAGE:4686857 5', mRNA sequence.
ACCESSION   BM171994
VERSION     BM171994.1  GI:17311557
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Kale,P.I., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and
            Prange,C.K.
  TITLE     The I.M.A.G.E. Consortium quality control effort: clone
            resequencing for verification
  JOURNAL   Unpublished
COMMENT     Other_ESTs: BG529561
            Contact: Prange CK
            The I.M.A.G.E. Consortium
            Lawrence Livermore National Laboratory
            Livermore, CA, USA
            Email: help@image.llnl.gov
            This read has been verified (found to hit its original self in the
            correct orientation), as part of the I.M.A.G.E. Consortium quality
            control effort. High quality sequence is defined as having 100 or
            more base pairs with a phred quality value of 20 or greater, where
            a sliding window of 4 base pairs with a phred quality value of 15
            or greater marks the beginning and end of the sequence. For
            information on obtaining this clone, please contact
            info@image.llnl.gov.
            Plate: LLCM1498  row: k  column: 18
            Seq primer: -21m13
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4686857"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_61"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT       74 a    146 c    117 g    116 t      1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 12;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        208 GGCTGGTGTCACCTGT 223


RESULT 35
BF035036/c
LOCUS       BF035036                455 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601456130F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3859734 5',
            mRNA sequence.
ACCESSION   BF035036
VERSION     BF035036.1  GI:10742748
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9593  row: l  column: 07
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3859734"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_66"
```

```
                     /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      129 a    117 c    133 g     76 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 10;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        138 GGCTGGTGTCACCTGT 123


Search completed: August 15, 2003, 10:57:39
Job time : 1136.8 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:32:41 ; Search time 492.975 Seconds
(without alignments)
1493.734 Million cell updates/sec

Title: US-10-074-620-1
Perfect score: 18
Sequence: 1 ggctggtgtcacctgtta 18

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	100.0	18	6	AX522236	AX522236 Sequence
c	2	18	100.0	175005	2	AC137657	AC137657 Sus scrof
c	3	17	94.4	88557	9	AL136314	AL136314 Human DNA
	4	17	94.4	107979	8	AP006146	AP006146 Lotus jap
c	5	17	94.4	226979	2	AC128140	AC128140 Rattus no
	6	17	94.4	231014	2	AC096132	AC096132 Rattus no
	7	17	94.4	238552	2	AC140198	AC140198 Mus muscu
	8	17	94.4	244424	2	AC131614	AC131614 Rattus no
c	9	17	94.4	244636	2	AC123215	AC123215 Rattus no
c	10	17	94.4	248793	2	AC103080	AC103080 Rattus no
	11	16	88.9	20	6	I26102	I26102 Sequence 28
c	12	16	88.9	50	6	AR032934	AR032934 Sequence
c	13	16	88.9	50	6	AR209598	AR209598 Sequence
c	14	16	88.9	50	6	I29674	I29674 Sequence 54
c	15	16	88.9	50	6	I91348	I91348 Sequence 54
c	16	16	88.9	435	6	BD025521	BD025521 Sequence
c	17	16	88.9	455	6	AX321046	AX321046 Sequence
c	18	16	88.9	963	9	BT006889	BT006889 Homo sapi
c	19	16	88.9	963	12	BT007572	BT007572 Synthetic
c	20	16	88.9	1041	6	I04174	I04174 Sequence 3
c	21	16	88.9	1103	9	AF020833	AF020833 Homo sapi
c	22	16	88.9	1115	9	HSU96074	U96074 Human trans
c	23	16	88.9	1128	9	BC000733	BC000733 Homo sapi
c	24	16	88.9	1138	9	BC008469	BC008469 Homo sapi
c	25	16	88.9	1142	6	BD063236	BD063236 Secreted
c	26	16	88.9	1174	9	AF094850	AF094850 Homo sapi
c	27	16	88.9	2199	8	AF348496	AF348496 Desmodesm
c	28	16	88.9	2556	14	AF172333	AF172333 Human her
c	29	16	88.9	2661	14	AF172332	AF172332 Human her
c	30	16	88.9	2661	14	EBVBLLF1A	X67776 Epstein-Bar
c	31	16	88.9	2663	14	EBVBLLF1	X99106 Epstein-Bar
c	32	16	88.9	2721	6	E00513	E00513 Genomic DNA
	33	16	88.9	3210	14	HS4GP340A	L07922 Epstein-Bar
c	34	16	88.9	3400	6	A11128	A11128 Synthetic n
c	35	16	88.9	3400	6	E01006	E01006 cDNA encodi
c	36	16	88.9	3833	6	AR049357	AR049357 Sequence
	37	16	88.9	4988	14	HS4GP340B	L07923 Epstein-Bar
c	38	16	88.9	5019	6	A11178	A11178 Synthetic n
c	39	16	88.9	5019	6	E01007	E01007 DNA sequenc
c	40	16	88.9	5931	6	AR233080	AR233080 Sequence
c	41	16	88.9	5931	14	HS4ENVGP	M10593 Epstein Bar
c	42	16	88.9	6450	9	AF092453	AF092453 Homo sapi
c	43	16	88.9	44904	5	TRU504410	AJ504410 Takifugu
	44	16	88.9	108036	9	AC084352	AC084352 Homo sapi
	45	16	88.9	128641	10	AL845521	AL845521 Mouse DNA
	46	16	88.9	129048	9	AC020931	AC020931 Homo sapi
	47	16	88.9	168012	2	AC025565	AC025565 Homo sapi
	48	16	88.9	169734	9	AC106748	AC106748 Homo sapi
	49	16	88.9	170803	2	AC016155	AC016155 Homo sapi
	50	16	88.9	171823	14	HHV507799	AJ507799 Human her
	51	16	88.9	172281	14	EBV	V01555 Epstein-Bar
	52	16	88.9	173950	9	AL390961	AL390961 Human DNA
c	53	16	88.9	174937	2	AC022245	AC022245 Homo sapi
	54	16	88.9	184113	14	HS4B958RAJ	M80517 Epstein-Bar
	55	16	88.9	194148	2	AC125349	AC125349 Mus muscu
	56	16	88.9	196560	2	AC132652	AC132652 Rattus no
	57	16	88.9	217717	2	BX530016	BX530016 Mus muscu
	58	16	88.9	225024	2	AC116415	AC116415 Mus muscu
c	59	16	88.9	234288	2	AL772311	AL772311 Mus muscu
c	60	16	88.9	244595	2	AC095144	AC095144 Rattus no
	61	16	88.9	244784	2	AC117356	AC117356 Rattus no
	62	16	88.9	247249	2	AC023526	AC023526 Homo sapi
c	63	16	88.9	266738	2	AL138896	AL138896 Homo sapi
c	64	15	83.3	266	11	G37013	G37013 SHGC-56377
c	65	15	83.3	622	11	BV025210	BV025210 S212P6604

c 66	15	83.3	688	10	AF327854	AF327854 Spermophi
67	15	83.3	781	11	BV059337	BV059337 S212P6928
68	15	83.3	1419	6	E54455	E54455 Gene causat
69	15	83.3	1419	10	AF124044	AF124044 Mus muscu
70	15	83.3	1808	6	AR262598	AR262598 Sequence
71	15	83.3	2293	6	E30803	E30803 Novel prote
72	15	83.3	2301	6	E30804	E30804 Novel prote
73	15	83.3	2952	6	AX017503	AX017503 Sequence
74	15	83.3	2952	6	BD135193	BD135193 Human nuc
75	15	83.3	3034	10	AK122327	AK122327 Mus muscu
76	15	83.3	3065	6	AR262597	AR262597 Sequence
c 77	15	83.3	3233	3	TBR308995	AJ308995 Trypanoso
78	15	83.3	3306	10	AB017609	AB017609 Mus muscu
79	15	83.3	3337	6	E30802	E30802 Novel prote
80	15	83.3	3660	10	AB017608	AB017608 Mus muscu
81	15	83.3	3674	6	E30801	E30801 Novel prote
82	15	83.3	3680	10	BC017126	BC017126 Mus muscu
c 83	15	83.3	4190	10	M60103	M60103 Rattus norv
c 84	15	83.3	4685	10	RNLAR1	X83505 R.norvegicu
c 85	15	83.3	6412	9	BC048768	BC048768 Homo sapi
c 86	15	83.3	6545	10	RATLARA	L11586 Rat leukocy
87	15	83.3	6550	6	AX453123	AX453123 Sequence
88	15	83.3	6553	6	AX453171	AX453171 Sequence
89	15	83.3	6553	8	AF332878	AF332878 Cochliobo
90	15	83.3	6672	6	AX335949	AX335949 Sequence
91	15	83.3	6672	9	HSU18259	U18259 Human clone
c 92	15	83.3	7702	6	AX658135	AX658135 Sequence
c 93	15	83.3	7702	9	HSLARR	Y00815 Human mRNA
c 94	15	83.3	10040	10	AF467766	AF467766 Mus muscu
95	15	83.3	14454	10	AB019041	AB019041 Mus muscu
c 96	15	83.3	33608	10	AC003057	AC003057 Mouse Cos
c 97	15	83.3	40438	9	AC087410	AC087410 Homo sapi
98	15	83.3	42115	6	AX453180	AX453180 Sequence
99	15	83.3	43101	3	AY190941	AY190941 Drosophil
100	15	83.3	43927	9	HSU69568	U69568 Human Xq28
101	15	83.3	51730	9	AF324889S2	AF324890 Homo sapi
c 102	15	83.3	53234	9	AC130889	AC130889 Homo sapi
c 103	15	83.3	57224	2	AC136542	AC136542 Rattus no
104	15	83.3	60761	2	AC102704	AC102704 Mus muscu
c 105	15	83.3	67197	9	AF510423S1	AF510423 Homo sapi
106	15	83.3	71983	9	AL445429	AL445429 Human DNA
107	15	83.3	72000	9	AP001102	AP001102 Homo sapi
108	15	83.3	73597	2	AC109598	AC109598 Homo sapi
109	15	83.3	74089	9	AL589762	AL589762 Human DNA
c 110	15	83.3	82419	9	AC004979	AC004979 Homo sapi
111	15	83.3	83549	9	AC004752	AC004752 Homo sapi
112	15	83.3	87157	9	AL136143	AL136143 Human DNA
c 113	15	83.3	88376	9	AL512407	AL512407 Human DNA
114	15	83.3	88554	10	AL669960	AL669960 Mouse DNA
c 115	15	83.3	90497	9	HS732E4	AL008722 Human DNA
116	15	83.3	101882	9	AC021089	AC021089 Homo sapi
117	15	83.3	101956	2	AP002494	AP002494 Homo sapi
118	15	83.3	108765	2	AF216674	AF216674 Homo sapi
c 119	15	83.3	109391	9	AL160031	AL160031 Human DNA
120	15	83.3	110000	2	AC106541_3	Continuation (4 of

ALIGNMENTS

```
RESULT 1
AX522236
LOCUS       AX522236                 18 bp    DNA     linear   PAT 24-OCT-2002
DEFINITION  Sequence 1 from Patent WO02064842.
ACCESSION   AX522236
VERSION     AX522236.1  GI:24411114
KEYWORDS    .
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1
  AUTHORS   Witte,D.P. and Groen,P.A.
  TITLE     Quantitative epstein barr virus pcr rapid assay
  JOURNAL   Patent: WO 02064842-A 1 22-AUG-2002;
            Children's Hospital Research Foundation (US)
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
BASE COUNT        2 a      4 c      6 g      6 t
ORIGIN

  Query Match           100.0%;  Score 18;  DB 6;  Length 18;
  Best Local Similarity  100.0%;  Pred. No. 0.81;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGTTA 18
              ||||||||||||||||||
Db          1 GGCTGGTGTCACCTGTTA 18


RESULT 2
AC137657/c
LOCUS       AC137657               175005 bp    DNA     linear   HTG 21-MAR-2003
DEFINITION  Sus scrofa clone RP44-494H16, WORKING DRAFT SEQUENCE, 12 ordered
            pieces.
ACCESSION   AC137657
VERSION     AC137657.2  GI:29135576
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 175005)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175005)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 175005)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Mar 21, 2003 this sequence version replaced gi:25700398.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ecq
                 Center clone name: 494H16

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
```

```
        coverage in Q20 bases and has been reviewed to rule out
        gross misassemblies, the low-quality ends of sequence
        contigs have been trimmed away, and each base is associated
        with a Phrap-derived quality score.
        -------------- Summary Statistics
             Sequencing vector: plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 171640 bases at least Q40
             Consensus quality: 172768 bases at least Q30
             Consensus quality: 173465 bases at least Q20
             Insert size: 168000; agarose-fp
             Insert size: 173905; sum-of-contigs
             Quality coverage: 9.32x in Q20 bases; agarose-fp
             Quality coverage: 9.01x in Q20 bases; sum-of-contigs
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 12 contigs. Gaps between the contigs
        * are represented as runs of N. The order of the pieces
        * is believed to be correct as given, however the sizes
        * of the gaps between them are based on estimates that have
        * provided by the submittor.
        * This sequence will be replaced
        * by the finished sequence as soon as it is available and
        * the accession number will be preserved.
        *        1     8074: contig of 8074 bp in length
        *     8075     8174: gap of unknown length
        *     8175    33811: contig of 25637 bp in length
        *    33812    33911: gap of unknown length
        *    33912    36676: contig of 2765 bp in length
        *    36677    36776: gap of unknown length
        *    36777    50073: contig of 13297 bp in length
        *    50074    50173: gap of unknown length
        *    50174    57164: contig of 6991 bp in length
        *    57165    57264: gap of unknown length
        *    57265    59445: contig of 2181 bp in length
        *    59446    59545: gap of unknown length
        *    59546    72773: contig of 13228 bp in length
        *    72774    72873: gap of unknown length
        *    72874   124522: contig of 51649 bp in length
        *   124523   124622: gap of unknown length
        *   124623   143805: contig of 19183 bp in length
        *   143806   143905: gap of unknown length
        *   143906   149152: contig of 5247 bp in length
        *   149153   149252: gap of unknown length
        *   149253   167607: contig of 18355 bp in length
        *   167608   167707: gap of unknown length
        *   167708   175005: contig of 7298 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175005
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9823"
                     /clone="RP44-494H16"
                     /clone_lib="RP44"
     misc_feature    1. .8074
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    8175. .33811
                     /note="assembly_fragment"
     misc_feature    33912. .36676
                     /note="assembly_fragment"
     misc_feature    36777. .50073
                     /note="assembly_fragment"
     misc_feature    50174. .57164
                     /note="assembly_fragment"
     misc_feature    57265. .59445
                     /note="assembly_fragment"
     misc_feature    59546. .72773
                     /note="assembly_fragment"
     misc_feature    72874. .124522
                     /note="assembly_fragment"
     misc_feature    124623. .143805
                     /note="assembly_fragment"
     misc_feature    143906. .149152
                     /note="assembly_fragment"
     misc_feature    149253. .167607
                     /note="assembly_fragment"
     misc_feature    167708. .175005
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    36244 a  47550 c  50242 g  39869 t   1100 others
ORIGIN

  Query Match             100.0%;  Score 18;  DB 2;  Length 175005;
  Best Local Similarity   100.0%;  Pred. No. 0.66;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGTTA 18
              ||||||||||||||||||
Db      18124 GGCTGGTGTCACCTGTTA 18107


RESULT 3
AL136314/c
LOCUS       AL136314               88557 bp    DNA     linear   PRI 01-NOV-2000
DEFINITION  Human DNA sequence from clone RP4-579A14 on chromosome 6, complete
            sequence.
ACCESSION   AL136314
VERSION     AL136314.12  GI:11120979
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 88557)
  AUTHORS   Peck,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10931919.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP4-579A14 is from the library RPCI-4 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP4-579A14 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-579A14 is at 1 in this sequence. The
            true left end of clone RP1-230L10 is at 88458 in this sequence. The
            true right end of clone RP5-826L7 is at 60956 in this sequence.
FEATURES             Location/Qualifiers
```

```
source          1. .88557
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="6"
                /clone="RP4-579A14"
                /clone_lib="RPCI-4"
repeat_region   778. .1087
                /note="AluY repeat: matches 1. .305 of consensus"
repeat_region   1795. .1850
                /note="14 copies 4 mer tata 83% conserved"
repeat_region   1853. .1884
                /note="16 copies 2 mer at 87% conserved"
repeat_region   1901. .2184
                /note="AluSg repeat: matches 1. .284 of consensus"
repeat_region   2380. .2991
                /note="36 copies 17 mer 80% conserved"
repeat_region   4022. .4061
                /note="MIR repeat: matches 204. .243 of consensus"
repeat_region   4062. .4247
                /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region   4248. .4357
                /note="MIR repeat: matches 94. .204 of consensus"
repeat_region   4544. .4676
                /note="7 copies 19 mer 88% conserved"
repeat_region   4680. .4831
                /note="8 copies 19 mer 89% conserved"
repeat_region   5642. .6058
                /note="MSTB repeat: matches 1. .426 of consensus"
misc_feature    6319. .6326
                /note="1318 bases of transposon Tn10 (J01829) removed
                here. This sequence represents the duplicated flanking
                sequence of the transposon."
repeat_region   7142. .7177
                /note="9 copies 4 mer acac 88% conserved"
repeat_region   7337. .7538
                /note="HAL1 repeat: matches 700. .906 of consensus"
repeat_region   8006. .8248
                /note="AluSx repeat: matches 40. .280 of consensus"
repeat_region   8253. .8684
                /note="MLT1C repeat: matches 22. .455 of consensus"
repeat_region   8732. .9214
                /note="L1PA13 repeat: matches 5663. .6149 of consensus"
repeat_region   10302. .10348
                /note="MER91A repeat: matches 120. .166 of consensus"
repeat_region   10349. .10724
                /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region   10732. .11084
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region   11085. .11195
                /note="MER91A repeat: matches 1. .114 of consensus"
repeat_region   11489. .11873
                /note="MER57B repeat: matches 1. .403 of consensus"
repeat_region   12045. .12287
                /note="AluSx repeat: matches 1. .242 of consensus"
repeat_region   12836. .13137
                /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region   13782. .14092
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   14290. .14600
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   14841. .15132
                /note="AluSg repeat: matches 1. .290 of consensus"
repeat_region   15665. .15962
                /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region   16080. .16559
                /note="L1MC/D repeat: matches 5319. .5572 of consensus"
repeat_region   16793. .17160
                /note="MER90 repeat: matches 176. .575 of consensus"
repeat_region   17161. .17285
                /note="MSTA repeat: matches 1. .130 of consensus"
repeat_region   17286. .17448
                /note="MER90 repeat: matches 32. .202 of consensus"
repeat_region   19375. .19574
                /note="100 copies 2 mer at 86% conserved"
repeat_region   19387. .19574
                /note="2 copies 94 mer 87% conserved"
repeat_region   19577. .19847
                /note="AluSx repeat: matches 1. .281 of consensus"
misc_feature    19798. .20271
                /note="match: GSS: Em:AQ338425"
repeat_region   20076. .20587
                /note="MLT1D repeat: matches 1. .498 of consensus"
misc_feature    21276. .22113
                /note="CpG island"
                /evidence=not_experimental
repeat_region   21445. .21527
                /note="MLT1A1 repeat: matches 405. .501 of consensus"
repeat_region   22372. .22666
                /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region   22796. .23027
                /note="L2 repeat: matches 2425. .2664 of consensus"
repeat_region   23896. .24202
                /note="MLT1F repeat: matches 228. .541 of consensus"
repeat_region   24203. .24563
                /note="THE1B repeat: matches 1. .362 of consensus"
repeat_region   24564. .24793
                /note="MLT1F repeat: matches 1. .228 of consensus"
repeat_region   26108. .26267
                /note="40 copies 4 mer tgtg 65% conserved"
repeat_region   26109. .26268
                /note="80 copies 2 mer gt 65% conserved"
repeat_region   26281. .26516
                /note="59 copies 4 mer tgtg 71% conserved"
repeat_region   26282. .26563
                /note="3 copies 94 mer 73% conserved"
repeat_region   26284. .26517
                /note="117 copies 2 mer gt 71% conserved"
repeat_region   27297. .27445
                /note="MIR repeat: matches 40. .190 of consensus"
repeat_region   28272. .28443
                /note="MIR repeat: matches 16. .208 of consensus"
repeat_region   29218. .29413
                /note="MER58A repeat: matches 1. .191 of consensus"
repeat_region   29407. .29520
                /note="6 copies 19 mer 68% conserved"
repeat_region   29414. .29497
                /note="21 copies 4 mer cttt 85% conserved"
repeat_region   29415. .29504
                /note="45 copies 2 mer tt 71% conserved"
repeat_region   29431. .29515
                /note="5 copies 17 mer 80% conserved"
repeat_region   29522. .29811
                /note="AluSq repeat: matches 7. .295 of consensus"
misc_feature    complement(29578. .30078)
                /note="match: GSS: Em:AQ894713"
repeat_region   29821. .29870
                /note="MER58A repeat: matches 175. .224 of consensus"
misc_feature    30089. .30632
                /note="match: GSS: Em:AQ318912"
repeat_region   32008. .32314
                /note="AluY repeat: matches 1. .306 of consensus"
repeat_region   32331. .32394
                /note="16 copies 4 mer agag 82% conserved"
repeat_region   32521. .32918
                /note="MLT1H repeat: matches 112. .532 of consensus"
repeat_region   33902. .34206
                /note="AluY repeat: matches 1. .305 of consensus"
repeat_region   34277. .34569
                /note="AluY repeat: matches 7. .297 of consensus"
repeat_region   34594. .34840
                /note="MLT1E repeat: matches 308. .568 of consensus"
repeat_region   35170. .35464
                /note="MLT1A1 repeat: matches 65. .365 of consensus"
repeat_region   35465. .35756
                /note="AluSc repeat: matches 4. .295 of consensus"
```

```
     repeat_region   35757. .35822
                     /note="MLT1A1 repeat: matches 1. .65 of consensus"
     repeat_region   37089. .37176
                     /note="MER5A repeat: matches 97. .187 of consensus"
     repeat_region   37728. .37832

  Query Match             94.4%;  Score 17;  DB 9;  Length 88557;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db      67152 GCTGGTGTCACCTGTTA 67136


RESULT 4
AP006146
LOCUS       AP006146               107979 bp    DNA     linear   PLN 19-MAR-2003
DEFINITION  Lotus japonicus genomic DNA, chromosome 3, clone:LjT34M22, TM0282,
            complete sequence.
ACCESSION   AP006146
VERSION     AP006146.1  GI:29122785
KEYWORDS    HTG.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Kato,T., Sato,S., Nakamura,Y. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. III. Sequence
            Features and Mapping of Sixty-two TAC Clones Which Cover the 6.7 Mb
            Regions of the Genome
  JOURNAL   DNA Res. 10, 27-33 (2003)
REFERENCE   2  (bases 1 to 107979)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .107979
                     /organism="Lotus japonicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:34305"
                     /chromosome="3"
                     /clone="LjT34M22"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0282"
BASE COUNT    33241 a  21293 c  19709 g  33736 t
ORIGIN

  Query Match             94.4%;  Score 17;  DB 8;  Length 107979;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db       1793 GCTGGTGTCACCTGTTA 1809


RESULT 5
AC128140/c
LOCUS       AC128140               226979 bp    DNA     linear   HTG 21-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-63O23, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC128140
VERSION     AC128140.2  GI:23265194
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 226979)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226979)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 226979)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 21, 2002 this sequence version replaced gi:21908739.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
```

```
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYAP
                Center clone name: CH230-63O23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 196911 bases at least Q40
                Consensus quality: 199136 bases at least Q30
                Consensus quality: 200537 bases at least Q20
                Estimated insert size: 209329; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   224410: contig of 224410 bp in length
            *   224411   224510: gap of unknown length
            *   224511   226979: contig of 2469 bp in length.
FEATURES             Location/Qualifiers
     source          1. .226979
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-63O23"
     misc_feature    977. .1828
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH321166"
     misc_feature    85339. .86666
                     /note="wgs_contig"
     misc_feature    complement(223202. .224034)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH321164"
BASE COUNT    54404 a  50080 c  47417 g  49761 t  25317 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 2;  Length 226979;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db     137713 GCTGGTGTCACCTGTTA 137697


RESULT 6
AC096132
LOCUS       AC096132               231014 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-11C18, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC096132
VERSION     AC096132.6  GI:30522660
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231014)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231014)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231014)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24818180.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
```

```
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GEMU
              Center clone name: CH230-11C18
          -------------- Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 208633 bases at least Q40
              Consensus quality: 212147 bases at least Q30
              Consensus quality: 214386 bases at least Q20
              Estimated insert size: 214070; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1   229193: contig of 229193 bp in length
          *  229194   229293: gap of unknown length
          *  229294   231014: contig of 1721 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231014
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-11C18"
     misc_feature    1. .1031
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2512. .3242
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH320811"
     misc_feature    23368. .24820
                     /note="wgs_contig"
     misc_feature    121675. .123616
                     /note="wgs_contig"
     misc_feature    124270. .126190
                     /note="wgs_contig"
     misc_feature    169881. .171027
                     /note="wgs_contig"
     misc_feature    complement(228041. .228684)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH320830"
BASE COUNT    61558 a  48903 c  47136 g  58505 t  14912 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 2;  Length 231014;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGTT 17
              |||||||||||||||||
Db     122048 GGCTGGTGTCACCTGTT 122064
```

```
RESULT 7
AC140198
LOCUS       AC140198              238552 bp    DNA     linear   HTG 01-MAR-2003
DEFINITION  Mus musculus chromosome UNK clone RP23-255C4, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC140198
VERSION     AC140198.2  GI:28626914
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 238552)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238552)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 238552)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 1, 2003 this sequence version replaced gi:28475433.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0255C04
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 236489 bases at least Q40
            Consensus quality: 236878 bases at least Q30
            Consensus quality: 237189 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 248898; sum-of-contigs
            Quality coverage: 16.05 in Q20 bases; agarose-fp
            Quality coverage: 12.12 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2229: contig of 2229 bp in length
            *    2230     2329: gap of unknown length
            *    2330    28154: contig of 25825 bp in length
            *   28155    28254: gap of unknown length
            *   28255    98428: contig of 70174 bp in length
            *   98429    98528: gap of unknown length
            *   98529   238552: contig of 140024 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238552
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-255C4"
```

```
     misc_feature    1. .2229
                     /note="assembly_name:Contig22"
     misc_feature    2330. .28154
                     /note="assembly_name:Contig24"
     misc_feature    28255. .98428
                     /note="assembly_name:Contig25"
     misc_feature    98529. .238552
                     /note="assembly_name:Contig26"
BASE COUNT    69322 a  53309 c  49714 g  65901 t    306 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 2;  Length 238552;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db     155734 GCTGGTGTCACCTGTTA 155750


RESULT 8
AC131614
LOCUS       AC131614               244424 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-284C9, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC131614
VERSION     AC131614.3  GI:25138981
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 244424)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244424)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244424)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23195985.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCCL
                Center clone name: CH230-284C9
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 221332 bases at least Q40
                Consensus quality: 224113 bases at least Q30
                Consensus quality: 225930 bases at least Q20
                Estimated insert size: 226659; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   235352: contig of 235352 bp in length
            *  235353   235452: gap of unknown length
            *  235453   240268: contig of 4816 bp in length
            *  240269   240368: gap of unknown length
            *  240369   244424: contig of 4056 bp in length.
FEATURES             Location/Qualifiers
     source          1. .244424
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
```

```
                     /clone="CH230-284C9"
     misc_feature    1. .1120
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1249. .2544
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(3470. .4259)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RXAHE17TJB"
BASE COUNT    56117 a  54386 c  56712 g  59935 t  17274 others
ORIGIN
```

Query Match 94.4%; Score 17; DB 2; Length 244424;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db     203097 GCTGGTGTCACCTGTTA 203113
```

```
RESULT 9
AC123215/c
LOCUS       AC123215               244636 bp    DNA     linear   HTG 21-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-118B7, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC123215
VERSION     AC123215.3  GI:23265875
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 244636)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 244636)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 244636)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 21, 2002 this sequence version replaced gi:21744493.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads. Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GSCP
                Center clone name: CH230-118B7
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 223663 bases at least Q40
                Consensus quality: 225326 bases at least Q30
                Consensus quality: 226426 bases at least Q20
                Estimated insert size: 241958; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   244636: contig of 244636 bp in length.
FEATURES            Location/Qualifiers
     source         1. .244636
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-118B7"
     misc_feature   1034. .1835
                    /note="clone_boundary
                    clone_end:Sp6
```

```
                     site:EcoRI
                     end_sequence:BH285202"
     misc_feature    complement(244009. .244636)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH285200"
BASE COUNT    62840 a  49128 c  49531 g  65556 t  17581 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 2;  Length 244636;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGTT 17
              |||||||||||||||||
Db      45298 GGCTGGTGTCACCTGTT 45282


RESULT 10
AC103080/c
LOCUS       AC103080              248793 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-225O1, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC103080
VERSION     AC103080.6  GI:30579747
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 248793)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 248793)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 248793)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23123516.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJEC
                Center clone name: CH230-225O1
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 213558 bases at least Q40
                Consensus quality: 215685 bases at least Q30
                Consensus quality: 217422 bases at least Q20
                Estimated insert size: 218947; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   243312: contig of 243312 bp in length
            *  243313   243412: gap of unknown length
            *  243413   244623: contig of 1211 bp in length
            *  244624   244723: gap of unknown length
            *  244724   245933: contig of 1210 bp in length
            *  245934   246033: gap of unknown length
            *  246034   247253: contig of 1220 bp in length
            *  247254   247353: gap of unknown length
            *  247354   248793: contig of 1440 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248793
```

```
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-225O1"
     misc_feature    1998. .2816
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BZ113888"
     misc_feature    complement(241014. .241843)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BZ113888"
BASE COUNT    59791 a  53488 c  51802 g  54201 t  29511 others
ORIGIN

  Query Match             94.4%;  Score 17;  DB 2;  Length 248793;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GCTGGTGTCACCTGTTA 18
              |||||||||||||||||
Db     183996 GCTGGTGTCACCTGTTA 183980


RESULT 11
I26102
LOCUS       I26102                    20 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 28 from patent US 5556772.
ACCESSION   I26102
VERSION     I26102.1  GI:1605972
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Sorge,J.A. and Mullinax,R.L.
  TITLE     Polymerase compositions and uses thereof
  JOURNAL   Patent: US 5556772-A 28 17-SEP-1996;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        2 a      4 c      7 g      7 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db          1 GGCTGGTGTCACCTGT 16


RESULT 12
AR032934/c
LOCUS       AR032934                  50 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 546 from patent US 5869241.
ACCESSION   AR032934
VERSION     AR032934.1  GI:5948539
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
  TITLE     Method of determining DNA sequence preference of a DNA-binding
            molecule
  JOURNAL   Patent: US 5869241-A 546 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="unknown"
BASE COUNT       18 a     14 c     10 g      8 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 13
AR209598/c
LOCUS       AR209598                  50 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 546 from patent US 6384208.
ACCESSION   AR209598
VERSION     AR209598.1  GI:21511065
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
  TITLE     Sequence directed DNA binding molecules compositions and methods
  JOURNAL   Patent: US 6384208-A 546 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="unknown"
BASE COUNT       18 a     14 c     10 g      8 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 14
I29674/c
LOCUS       I29674                    50 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 546 from patent US 5578444.
ACCESSION   I29674
VERSION     I29674.1  GI:1820465
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.
  TITLE     Sequence-directed DNA-binding molecules compositions and methods
  JOURNAL   Patent: US 5578444-A 546 26-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="unknown"
BASE COUNT       18 a     14 c     10 g      8 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 15
I91348/c
LOCUS       I91348                    50 bp    DNA     linear   PAT 01-DEC-1998
```

```
DEFINITION  Sequence 546 from patent US 5726014.
ACCESSION   I91348
VERSION     I91348.1  GI:3935818
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.
  TITLE     Screening assay for the detection of DNA-binding molecules
  JOURNAL   Patent: US 5726014-A 546 10-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="unknown"
BASE COUNT       18 a     14 c     10 g      8 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db         41 GGCTGGTGTCACCTGT 26


RESULT 16
BD025521/c
LOCUS       BD025521                435 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Sequence tag and encoded human protein.
ACCESSION   BD025521
VERSION     BD025521.1  GI:22566744
KEYWORDS    JP 2001269182-A/1767.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
  TITLE     Sequence tag and encoded human protein
  JOURNAL   Patent: JP 2001269182-A 1767 02-OCT-2001;
            GENSET
COMMENT     OS   Homo sapiens (human)
            PN   JP 2001269182-A/1767
            PD   02-OCT-2001
            PF   24-FEB-2000 JP 2000118773
            PR   26-FEB-1999 US    60/122487
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
            PI   JORDAN
            PC   C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
             C12N5/10,
            PC   C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
              G06F15/40
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             60. .434.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      105 a    114 c    133 g     79 t      4 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        186 GGCTGGTGTCACCTGT 171


RESULT 17
AX321046/c
LOCUS       AX321046                455 bp    DNA     linear   PAT 15-DEC-2001
DEFINITION  Sequence 63 from Patent WO0177168.
ACCESSION   AX321046
VERSION     AX321046.1  GI:17904325
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0177168-A 63 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      122 a    116 c    138 g     78 t      1 others
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        177 GGCTGGTGTCACCTGT 162


RESULT 18
BT006889/c
LOCUS       BT006889                963 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens eukaryotic translation initiation factor 3, subunit 4
            delta, 44kDa mRNA, complete cds.
ACCESSION   BT006889
VERSION     BT006889.1  GI:30582616
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 963)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 963)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .963
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00457X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .963
                     /codon_start=1
                     /product="eukaryotic translation initiation factor 3,
                     subunit 4 delta, 44kDa"
                     /protein_id="AAP35535.1"
                     /db_xref="GI:30582617"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      237 a    291 c    288 g    147 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 963;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        127 GGCTGGTGTCACCTGT 112


RESULT 19
BT007572/c
LOCUS       BT007572                963 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens eukaryotic translation initiation
            factor 3, subunit 4 delta, 44kDa mRNA, partial cds.
ACCESSION   BT007572
VERSION     BT007572.1  GI:30583982
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 963)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 963)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .963
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00457L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>963
                     /note="Mutations: 962:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens eukaryotic translation initiation
                     factor 3, subunit 4 delta, 44kDa"
                     /protein_id="AAP36240.1"
                     /db_xref="GI:30583983"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTNL"
BASE COUNT      236 a    291 c    288 g    148 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 12;  Length 963;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        127 GGCTGGTGTCACCTGT 112


RESULT 20
I04174/c
LOCUS       I04174                  1041 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 3 from Patent US 4707358.
ACCESSION   I04174
VERSION     I04174.1  GI:268757
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Kieff,E., Tanner,J., Hummel,M. and Beisel,C.
  TITLE     Vaccine against Epstein-Barr Virus
  JOURNAL   Patent: US 4707358-A 3 17-NOV-1987;
            The University of Chicago; Chicago, IL
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      295 a    383 c    199 g    164 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        444 GGCTGGTGTCACCTGT 429


RESULT 21
AF020833/c
LOCUS       AF020833                1103 bp    mRNA    linear   PRI 02-MAR-1999
DEFINITION  Homo sapiens eukaryotic translation initiation factor 3 subunit
            (p42) mRNA, complete cds.
ACCESSION   AF020833
VERSION     AF020833.1  GI:2460199
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1103)
  AUTHORS   Bandyopadhyay,A. and Maitra,U.
```

```
  TITLE     Cloning and characterization of the p42 subunit of mammalian
            translation initiation factor 3 (eIF3): demonstration that eIF3
            interacts with eIF5 in mammalian cells
  JOURNAL   Nucleic Acids Res. 27 (5), 1331-1337 (1999)
  MEDLINE   99141230
   PUBMED   9973622
REFERENCE   2  (bases 1 to 1103)
  AUTHORS   Bandyopadhyay,A., Chaudhuri,J., Si,K., Tempst,P. and Maitra,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1997) Dev. & Molecular Biol., Albert Einstein
            College of Medicine of Yeshiva University, 1300, Morris Park
            Avenue, Bronx, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .1103
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
     gene            1. .1103
                     /gene="p42"
     CDS             24. .986
                     /gene="p42"
                     /note="Similar to WP:F22B5.2 CE02197 RNA binding protein"
                     /codon_start=1
                     /product="eukaryotic translation initiation factor 3
                     subunit"
                     /protein_id="AAB71866.1"
                     /db_xref="GI:2460200"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      280 a    332 c    324 g    167 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 1103;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        150 GGCTGGTGTCACCTGT 135


RESULT 22
HSU96074/c
LOCUS       HSU96074                1115 bp    mRNA    linear   PRI 24-NOV-1998
DEFINITION  Human translation initiation factor eIF3 p44 subunit mRNA, complete
            cds.
ACCESSION   U96074
VERSION     U96074.1  GI:3264858
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1115)
  AUTHORS   Block,K.L., Vornlocher,H.P. and Hershey,J.W.
  TITLE     Characterization of cDNAs encoding the p44 and p35 subunits of
            human translation initiation factor eIF3
  JOURNAL   J. Biol. Chem. 273 (48), 31901-31908 (1998)
  MEDLINE   99041954
   PUBMED   9822659
REFERENCE   2  (bases 1 to 1115)
  AUTHORS   Block,K.L., Vornlocher,H.-P. and Hershey,J.W.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1997) Biological Chemistry, U C Davis, Tupper
            Hall, Davis, CA 95616, USA
FEATURES             Location/Qualifiers
     source          1. .1115
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             10. .972
                     /codon_start=1
                     /product="translation initiation factor eIF3 p44 subunit"
                     /protein_id="AAC78728.1"
                     /db_xref="GI:3264859"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDRARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      305 a    328 c    318 g    164 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 1115;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        136 GGCTGGTGTCACCTGT 121


RESULT 23
BC000733/c
LOCUS       BC000733                1128 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, eukaryotic translation initiation factor 3, subunit 4
            (delta, 44kD), clone MGC:2053 IMAGE:3504640, mRNA, complete cds.
ACCESSION   BC000733
VERSION     BC000733.1  GI:12653882
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1128)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: e Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 10280561.
FEATURES             Location/Qualifiers
     source          1. .1128
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:8666"
                     /db_xref="taxon:9606"
                     /clone="MGC:2053 IMAGE:3504640"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             52. .1014
```

```
                     /codon_start=1
                     /product="eukaryotic translation initiation factor 3,
                     subunit 4 (delta, 44kD)"
                     /protein_id="AAH00733.1"
                     /db_xref="GI:12653883"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      283 a    336 c    335 g    174 t
ORIGIN

  Query Match          88.9%;  Score 16;  DB 9;  Length 1128;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        178 GGCTGGTGTCACCTGT 163


RESULT 24
BC008469/c
LOCUS       BC008469               1138 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, eukaryotic translation initiation factor 3, subunit 4
            (delta, 44kD), clone MGC:14741 IMAGE:4279770, mRNA, complete cds.
ACCESSION   BC008469
VERSION     BC008469.1  GI:14250113
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1138)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 21 Row: e Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .1138
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:8666"
                     /db_xref="taxon:9606"
                     /clone="MGC:14741 IMAGE:4279770"
                     /tissue_type="Brain, primitive neuroectodermal"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             44. .1006
                     /codon_start=1
                     /product="eukaryotic translation initiation factor 3,
                     subunit 4 (delta, 44kD)"
                     /protein_id="AAH08469.1"
                     /db_xref="GI:14250114"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      297 a    334 c    331 g    176 t
ORIGIN

  Query Match          88.9%;  Score 16;  DB 9;  Length 1138;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        170 GGCTGGTGTCACCTGT 155


RESULT 25
BD063236/c
LOCUS       BD063236               1142 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Secreted human proteins.
ACCESSION   BD063236
VERSION     BD063236.1  GI:22608839
KEYWORDS    JP 2001505783-A/11.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1142)
  AUTHORS   Escobedo,J., Hu,Q., Garcia,P., Williams,L.T. and Kothakota,S.
  TITLE     Secreted human proteins
  JOURNAL   Patent: JP 2001505783-A 11 08-MAY-2001;
            CHIRON CORP
COMMENT     PN   JP 2001505783-A/11
            PD   08-MAY-2001
            PF   11-DEC-1997 JP 1998526977
            PR   11-DEC-1996 US    60/032757
            PI   JAIME ESCOBEDO,QUIANJIN HU,PABLO GARCIA,LEWIS T WILLIAMS PI
            ,SRINIVAS KOTHAKOTA
            PC   C12N15/12,C12N15/62,C12N15/85,C12N5/10,C12N1/21,C07K14/47, PC
             C07K16/18
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1142
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
BASE COUNT      304 a    335 c    331 g    172 t
ORIGIN

  Query Match          88.9%;  Score 16;  DB 6;  Length 1142;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        159 GGCTGGTGTCACCTGT 144


RESULT 26
AF094850/c
LOCUS       AF094850               1174 bp    mRNA    linear   PRI 23-MAR-2001
DEFINITION  Homo sapiens eukaryotic translation initiation factor 3 subunit
            p42/p44 mRNA, complete cds.
ACCESSION   AF094850
VERSION     AF094850.1  GI:10280561
```

```
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1174)
  AUTHORS   Chen,W., Blough,R.I. and Winkelmann,J.C.
  TITLE     Molecular cloning, genomic structure and chromosomal localization
            of a novel human RNA binding protein gene homologous to a tumor
            necrosis factor alpha inducible transcript in mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1174)
  AUTHORS   Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppes,B. and
            Winkelmann,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology,
            University of Cincinnati College of Medicine, 231 Bethesda Ave.,
            Cincinnati, OH 45267-0508, USA
FEATURES             Location/Qualifiers
     source          1. .1174
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.2"
     CDS             42. .1004
                     /note="eIF3 p42/p44"
                     /codon_start=1
                     /product="eukaryotic translation initiation factor 3
                     subunit p42/p44"
                     /protein_id="AAG15419.1"
                     /db_xref="GI:10280562"
                     /translation="MPTGDFDSKPSWADQVEEEGEDDKCVTSELLKGIPLATGDTSPE
                     PELLPGAPLPPPKEVINGNIKTVTEYKIDEDGKKFKIVRTFRIETRKASKAVARRKNW
                     KKFGNSEFDPPGPNVATTTVSDDVSMTFITSKEDLNCQEEEDPMNKLKGQKIVSCRIC
                     KGDHWTTRCPYKDTLGPMQKELAEQLGLSTGEKEKLPGELEPVQATQNKTGKYVPPSL
                     RDGASRRGESMQPNRRADDNATIRVTNLSEDTRETDLQELFRPFGSISRIYLAKDKTT
                     GQSKGFAFISFHRREDAARAIAGVSGFGYDHLILNVEWAKPSTN"
BASE COUNT      334 a    336 c    331 g    173 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 9;  Length 1174;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        168 GGCTGGTGTCACCTGT 153


RESULT 27
AF348496/c
LOCUS       AF348496               2199 bp    DNA     linear   PLN 21-AUG-2001
DEFINITION  Desmodesmus pirkollei 18S ribosomal RNA gene, complete sequence.
ACCESSION   AF348496
VERSION     AF348496.1  GI:15216663
KEYWORDS    .
SOURCE      Desmodesmus pirkollei
  ORGANISM  Desmodesmus pirkollei
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
            Chlorococcales; Scenedesmaceae; Desmodesmus.
REFERENCE   1  (bases 1 to 2199)
  AUTHORS   Hegewald,E., Coesel,P. and Hegewald,P.
  TITLE     A phytoplankton collection from Bali, with description of a new
            Desmodesmus species (Chlorophyta, Scenedesmaceae)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2199)
  AUTHORS   Hegewald,E., Coesel,P. and Hegewald,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2001) Institute of Chemistry and Dynamics of the
            Geosphere 6, Research Center Juelich, Juelich D-52425, Germany
FEATURES             Location/Qualifiers
     source          1. .2199
                     /organism="Desmodesmus pirkollei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:165818"
     rRNA            1. .2199
                     /product="18S ribosomal RNA"
BASE COUNT      572 a    465 c    593 g    569 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 8;  Length 2199;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       1393 GGCTGGTGTCACCTGT 1378


RESULT 28
AF172333/c
LOCUS       AF172333               2556 bp    mRNA    linear   VRL 29-AUG-1999
DEFINITION  Human herpesvirus 4 cell-line SNU-1103 major outer envelope
            glycoprotein gp350 mRNA, complete cds.
ACCESSION   AF172333
VERSION     AF172333.1  GI:5802484
KEYWORDS    .
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 2556)
  AUTHORS   Lee,W.K., Kim,S.M., Sim,Y.S., Cho,S.G., Park,S.H., Kim,C.W. and
            Park,J.G.
  TITLE     B-lymphoblastoid cell lines from cancer patients
  JOURNAL   In Vitro Cell. Dev. Biol. Anim. 34 (2), 97-100 (1998)
  MEDLINE   98203772
   PUBMED   9542645
REFERENCE   2  (bases 1 to 2556)
  AUTHORS   Chang,S.H., Kim,S.H., Lee,W.K., Kim,H.J., Choi,S.H., Park,J.H.,
            Jang,H.S., Chung,G.H., Kwon,T.H., Kim,D.H., Yang,M.S. and Jang,Y.S.
  TITLE     Cloning and analysis of the Epstein-Barr virus glycoprotein 350
            genes
  JOURNAL   Mol. Cells 8 (5), 585-593 (1998)
  MEDLINE   99072166
   PUBMED   9856346
REFERENCE   3  (bases 1 to 2556)
  AUTHORS   Chang,S.H., Kim,S.H., Lee,W.K., Kim,H.J., Choi,S.H., Park,J.H.,
            Jang,H.S., Chung,G.H., Kwon,T.H., Kim,D.H., Yang,M.S. and Jang,Y.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-1999) Faculty of Biological Sciences, Chonbuk
            National University, 664-14, Dukjin-dong 1-Ka, Dukjin-Ku, Chonju
            561 - 756, Korea
FEATURES             Location/Qualifiers
     source          1. .2556
                     /organism="Human herpesvirus 4"
                     /mol_type="mRNA"
                     /db_xref="taxon:10376"
                     /cell_line="SNU-1103"
     CDS             1. .2556
                     /codon_start=1
                     /product="major outer envelope glycoprotein gp350"
                     /protein_id="AAD51698.1"
                     /db_xref="GI:5802485"
                     /translation="MEAALLVCQYTIQSLIQLTRDDPGFFMVEILEFPFYPACNVCTA
                     DVNATINFDDGGKKHQLNLDFGQLTPHTMAVYQPRGAFGGSENATNLFRLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTQMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPMTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFAAPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPRDNGTESKAPDMTS
                     PTSAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTVGETSPQ
                     ANTTNHTLGGTSSTPVVTSPPKNATSAVTTGQHNITSSSTSSMSLRPSSISETLSPST
```

SDNSTSHMPLLTSAHPTGGENITQVTPASTSTHHVSTSSPAPRPGTTCQASGPGNSST
STKPGEVNVTKGTPPKNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTS
TEPTTDYGGDSTTPRTRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQP
RFSNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"

BASE COUNT 713 a 798 c 532 g 513 t
ORIGIN

Query Match 88.9%; Score 16; DB 14; Length 2556;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 1956 GGCTGGTGTCACCTGT 1941

RESULT 29
AF172332/c
LOCUS AF172332 2661 bp mRNA linear VRL 29-AUG-1999
DEFINITION Human herpesvirus 4 cell-line SNU-20 major outer envelope glycoprotein gp350 mRNA, complete cds.
ACCESSION AF172332
VERSION AF172332.1 GI:5802482
KEYWORDS .
SOURCE Human herpesvirus 4 (Epstein-Barr virus)
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 2661)
AUTHORS Lee,W.K., Kim,S.M., Sim,Y.S., Cho,S.G., Park,S.H., Kim,C.W. and Park,J.G.
TITLE B-lymphoblastoid cell lines from cancer patients
JOURNAL In Vitro Cell. Dev. Biol. Anim. 34 (2), 97-100 (1998)
MEDLINE 98203772
PUBMED 9542645
REFERENCE 2 (bases 1 to 2661)
AUTHORS Chang,S.H., Kim,S.H., Lee,W.K., Kim,H.J., Choi,S.H., Park,J.H., Jang,H.S., Chung,G.H., Kwon,T.H., Kim,D.H., Yang,M.S. and Jang,Y.S.
TITLE Cloning and analysis of the Epstein-Barr virus glycoprotein 350 genes
JOURNAL Mol. Cells 8 (5), 585-593 (1998)
MEDLINE 99072166
PUBMED 9856346
REFERENCE 3 (bases 1 to 2661)
AUTHORS Chang,S.H., Kim,S.H., Lee,W.K., Kim,H.J., Choi,S.H., Park,J.H., Jang,H.S., Chung,G.H., Kwon,T.H., Kim,D.H., Yang,M.S. and Jang,Y.S.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-1999) Faculty of Biological Sciences, Chonbuk National University, 664-14, Dukjin-dong 1-Ka, Dukjin-Ku, Chonju 561 - 756, Korea
FEATURES Location/Qualifiers
source 1. .2661
/organism="Human herpesvirus 4"
/mol_type="mRNA"
/db_xref="taxon:10376"
/cell_line="SNU-20"
CDS 1. .2661
/codon_start=1
/product="major outer envelope glycoprotein gp350"
/protein_id="AAD51697.1"
/db_xref="GI:5802483"
/translation="MEAALLVCQYTIQSLIQLTRDDPGFFNVEILEFPFYPACNVCTA
DVNATINFDVGGKKHKLNLDFGLLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINITAGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFAAPNTTTGLPS
STHVPTNLTAPTSTGPTVSTADVTSPTPAGTTSGASPVTPRPSPRDNGTESKAPDMTS
PTSAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPT
PNATIPTLGKTSPTSAVTTPTPNATSPTVGETSPKANTTNHTLGGTSSTPVVTSPPKN
ATSAVTTGQHNITSSSTSSMSLRPSSISETLSPSTSDNSMSHMPLLTSAHPTGGENIT
QVTPASTSTHHVSTSSPAPRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPKNATSPQA
PSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRTRYNATTY
LPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNLSMLVLQWASLAVLTLLLL
LVMADCAFRRNLSTSHTYTTPPYDDAETYV"
BASE COUNT 752 a 844 c 542 g 523 t
ORIGIN

Query Match 88.9%; Score 16; DB 14; Length 2661;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCTGGTGTCACCTGT 16
||||||||||||||||
Db 2061 GGCTGGTGTCACCTGT 2046

RESULT 30
EBVBLLF1A/c
LOCUS EBVBLLF1A 2661 bp DNA linear VRL 23-OCT-1992
DEFINITION Epstein-Barr virus BLLF1 gene for glycoprotein 350/220.
ACCESSION X67776
VERSION X67776.1 GI:59163
KEYWORDS glycoprotein 350/220.
SOURCE Human herpesvirus 4 (Epstein-Barr virus)
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 2661)
AUTHORS Klein,K. and Mueller-Lantzsch,N.
TITLE Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr virus type-B (P3HR1)
JOURNAL Nucleic Acids Res.
REFERENCE 2 (bases 1 to 2661)
AUTHORS Klein,K.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1992) K. Klein, Inst f Med Mikrobiologie u Hygiene, Abteilung Virologie, Universitaetskliniken des Saarlandes, 6650 Homburg/Saar, FRG
FEATURES Location/Qualifiers
source 1. .2661
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/strain="P3HR1"
/db_xref="taxon:10376"
gene 1. .2661
/gene="BLLF1"
CDS 1. .2661
/gene="BLLF1"
/codon_start=1
/product="glycoprotein 350/220"
/protein_id="CAA47986.1"
/db_xref="GI:59164"
/db_xref="SWISS-PROT:Q07284"
/translation="MEAALLVCQYTIQSLIQLTRDDPGFFNVEILEFPFYPACNVCTA
DVNATINFDVGGKKHKLNLDFGLLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINITTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFAAPNTTTGLPS
STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPRDNGTESKAPDMTS
PTSAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPT
PNATIPTLGKTSPTSAVTTPTPNATSPTVGETSPQANTTNHTLGGTSSTPVVTSPPKN
ATSAVTTGQHNITSSSTSSMSLRPSSISETLSPSTSDNSTSHMPLLTSAHPTGGENIT
QVTPASTSTHHVSTSSPAPRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPKNATSPQA
PSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRTRYNATTY
LPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNLSMLVLQWASLAVLTLLLL
LVMADCAFRRNLSTSHTYTTPPYDDAETYV"
BASE COUNT 750 a 847 c 542 g 522 t
ORIGIN

Query Match 88.9%; Score 16; DB 14; Length 2661;

```
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       2061 GGCTGGTGTCACCTGT 2046


RESULT 31
EBVBLLF1/c
LOCUS       EBVBLLF1                2663 bp    DNA     linear   VRL 09-JAN-1998
DEFINITION  Epstein-Barr virus BLLF1 gene..
ACCESSION   X99106
VERSION     X99106.1  GI:2769559
KEYWORDS    BLLF1 gene; gp340.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1
  AUTHORS   Mackett,M., Cox,C., Pepper,S.D., Lees,J.F., Naylor,B.A.,
            Wedderburn,N. and Arrand,J.R.
  TITLE     Immunisation of common marmosets with vaccinia virus expressing
            Epstein-Barr virus (EBV) gp340 and challenge with EBV
  JOURNAL   J. Med. Virol. 50 (3), 263-271 (1996)
  MEDLINE   97082049
   PUBMED   8923292
REFERENCE   2  (bases 1 to 2663)
  AUTHORS   Pepper,S.D.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) S.D.V. Pepper, Paterson Institute for
            Cancer Research, Molecular Biology, Wilmslow Rd, Withington,
            Manchester M20 9BX, UK
FEATURES             Location/Qualifiers
     source          1. .2663
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /strain="M81"
                     /db_xref="taxon:10376"
     gene            66. .2636
                     /gene="BLLF1"
     CDS             66. .2636
                     /gene="BLLF1"
                     /codon_start=1
                     /product="gp340"
                     /protein_id="CAA67558.1"
                     /db_xref="GI:2769560"
                     /db_xref="SPTREMBL:O56854"
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTQMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPRDNGTESKAPDMTS
                     PTSAVTTPTPNGTSPTPAMTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPT
                     PNATSPTVGETSPQANATNHTLGGTSPTPVVTSPPKNATSDVTTGQHNRTSSSTSSMS
                     LRPSSIPETTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGP
                     GNSSTSTKPGEVNVTKGTPPKNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGH
                     GARTSTEPTTDYGDDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVP
                     PTSQPRFSNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV
                     "
     polyA_signal    2641. .2646
BASE COUNT      755 a    829 c    552 g    527 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 2663;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
```

```
              ||||||||||||||||
Db       2036 GGCTGGTGTCACCTGT 2021


RESULT 32
E00513/c
LOCUS       E00513                  2721 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Genomic DNA encoding Epstein-Barr virus gp 220/200.
ACCESSION   E00513
VERSION     E00513.1  GI:2168792
KEYWORDS    JP 1985232094-A/1.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 2721)
  AUTHORS   Eriotsuto,K., Jieroomu,T., Mearii,F. and Kurisutofuaa,B.
  TITLE     VACCINE TO EPSTAIN VAR VIRUS
  JOURNAL   Patent: JP 1985232094-A 1 18-NOV-1985;
            UNIV CHICAGO:THE
COMMENT     OS   Epstein-Barr virus
            PN   JP 1985232094-A/1
            PD   18-NOV-1985
            PF   30-JAN-1985 JP 1985014606
            PR   30-JAN-1984 US 84    575352, 23-JUL-1984 US 84    633558 PI
            ERIOTSUTO KIIFU, JIEROOMU TANNAA, MEARII FUNMERU, PI   KURISUTOFUAA
            BEISERU
            PC   C12N15/00,A61K39/245,C07H21/04,C07K13/00,C12N1/00//C12P21/02,
            PC   (C12N1/00,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .>2721
            FT                   /product='gp 220/200'.
FEATURES             Location/Qualifiers
     source          1. .2721
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
BASE COUNT      762 a    876 c    557 g    526 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 2721;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       2124 GGCTGGTGTCACCTGT 2109


RESULT 33
HS4GP340A
LOCUS       HS4GP340A               3210 bp    DNA     linear   VRL 15-FEB-2001
DEFINITION  Epstein-Barr virus glycoprotein 340 (BLLF1) and BLLF2 genes,
            complete cds's, and BLLFR3 gene, first exon.
ACCESSION   L07922
VERSION     L07922.1  GI:291519
KEYWORDS    antigen; glycoprotein 340; gp340; viral antigen.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 3210)
  AUTHORS   Lees,J.F., Arrand,J.E., Pepper,S.D., Stewart,J.P., Mackett,M. and
            Arrand,J.R.
  TITLE     The Epstein-Barr virus candidate vaccine antigen gp340/220 is
            highly conserved between virus types A and B
  JOURNAL   Virology 195 (2), 578-586 (1993)
```

```
  MEDLINE   93331716
   PUBMED   8393237
FEATURES             Location/Qualifiers
     source          1. .3210
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /strain="P3HR-1"
                     /db_xref="taxon:10376"
     gene            complement(1. .2661)
                     /gene="BLLF1"
     CDS             complement(1. .2661)
                     /gene="BLLF1"
                     /codon_start=1
                     /product="glycoprotein 340"
                     /protein_id="AAA02783.1"
                     /db_xref="GI:291520"
                     /translation="MEAALLVCQYTIQSLIQLTRDDPGFFNVEILEFPFYPACNVCTA
                     DVNATINFDVGGKKHKLNLDFGLLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINITTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFAAPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPRDNGTESKAPDMTS
                     PTSAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPT
                     PNATIPTLGKTSPTSAVTTPTPNATSPTVGETSPQANTTNHTLGGTSSTPVVTSPPKN
                     ATSAVTTGQHNITSSSTSSMSLRPSSISETLSPSTSDNSTSHMPLLTSAHPTGGENIT
                     QVTPASTSTHHVSTSSPAPRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPKNATSPQA
                     PSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRTRYNATTY
                     LPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNLSMLVLQWASLAVLTLLLL
                     LVMADCAFRRNLSTSHTYTTPPYDDAETYV"
     gene            complement(138. .584)
                     /gene="BLLF2"
     CDS             complement(138. .584)
                     /gene="BLLF2"
                     /codon_start=1
                     /protein_id="AAA02784.1"
                     /db_xref="GI:291521"
                     /translation="MCPPVRQRPAQAPPAKRQALETVPHPQNRGRLMSPKARPPKMQR
                     RPRPPVAKRRRFPRSPQQVERPILPPVESTPQDMEPGQVQSPPQITAVIQLRQERDTM
                     RPPIYLPALLANCGPAGLLRAHRLPQPKPPCLSRQRPSPDSQTSPC"
     gene            2751. .3210
                     /gene="BLRF3"
     CDS             2751. .>3210
                     /gene="BLRF3"
                     /codon_start=1
                     /protein_id="AAA02785.2"
                     /db_xref="GI:12831528"
                     /translation="MDKDRPGLPAPDDNIEEVPSTSGVQERASEGDWENVLIEISDSS
                     SEEEADAHLDSSQRGKKRKRVDDDAGGSAPAQHVPPPQLDHPGREAILYRFPLDLRR
                     FIQAIGAAATVSFPMAQVCDVCFCPSHNKVSDLLPLVSAPRHASHRPVFRI"
     exon            2751. .3210
                     /gene="BLRF3"
                     /number=1
BASE COUNT      659 a    678 c   1001 g    872 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 14;  Length 3210;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db        601 GGCTGGTGTCACCTGT 616


RESULT 34
A11128/c
LOCUS       A11128                  3400 bp    DNA     linear   PAT 07-DEC-1993
DEFINITION  Synthetic nucleotide sequence of the leftward reading frame of the
            Bam L-fragment encoding gp 250/350.
ACCESSION   A11128
VERSION     A11128.1  GI:490970
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 3400)
  AUTHORS   Wolf,H.J.
  TITLE     DNA sequences of the EBV genome, recombinant DNA molecules,
            processes for producing EBV-related antigens, diagnostic
            compositions and pharmaceutical compositions containing said
            antigens
  JOURNAL   Patent: EP 0173254-A 3 05-MAR-1986;
            Wolf, Hans Joachim
FEATURES             Location/Qualifiers
     source          1. .3400
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
     CDS             556. .3279
                     /note="Leftward reading frame of the Bam L-fragment
                     encoding gp 250/350."
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00938.1"
                     /db_xref="GI:490971"
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                     STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                     PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                     ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                     NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                     KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                     PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                     SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
BASE COUNT      904 a   1060 c    732 g    704 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 3400;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       2679 GGCTGGTGTCACCTGT 2664


RESULT 35
E01006/c
LOCUS       E01006                  3400 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding p350 of Epstein-Barr virus.
ACCESSION   E01006
VERSION     E01006.1  GI:2169265
KEYWORDS    JP 1986257188-A/2.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 3400)
  AUTHORS   Hansu,Y.B.
  TITLE     DNA SEQUENCE OF EBV GENOM, RECOMBINED DNA MOLECULE AND EBV RELATED
            ANTIGEN AND DIAGNOSTIC COMPOSITION CONTAINING SAID ANTIGEN AND
            PRODUCTION OF PREPARATION
  JOURNAL   Patent: JP 1986257188-A 2 14-NOV-1986;
            HANSU YOTSUTO BUORUFU
COMMENT     OS   Epstein-Barr virus
            PN   JP 1986257188-A/2
```

```
            PD   14-NOV-1986
            PF   23-AUG-1985 JP 1985185661
            PR   23-AUG-1984 EP 84  84110089, 23-AUG-1984 EP 84  84110090 PI
            HANSU YOTSUTO BUORUFU
            PC   C12N15/00,A61K39/245,C07H21/04,C07K13/00,C12N1/00,G01N33/569,
            PC   G01N33/577//
            PC   C12P21/02,(C12N1/00,C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           1. .555
            FT   CDS             556. .3279
            FT                   /product='p350 Epstein-Barr virus' FT   3'UTR
                 3280. .3400
            FT   polyA_site      3284. .3291.
FEATURES             Location/Qualifiers
     source          1. .3400
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
BASE COUNT      904 a   1060 c    732 g    704 t
ORIGIN

  Query Match             88.9%;  Score 16;  DB 6;  Length 3400;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCTGGTGTCACCTGT 16
              ||||||||||||||||
Db       2679 GGCTGGTGTCACCTGT 2664


Search completed: August 15, 2003, 09:33:48
Job time : 503.975 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:32:41 ; Search time 547.75 Seconds
(without alignments)
1493.734 Million cell updates/sec

Title: US-10-074-620-2
Perfect score: 20
Sequence: 1 ccttaggaggaacaagtccc 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	100.0	20	6	AX522237	AX522237 Sequence
	2	20	100.0	20	6	I26103	I26103 Sequence 29
	3	20	100.0	1041	6	I04174	I04174 Sequence 3
	4	20	100.0	2663	14	EBVBLLF1	X99106 Epstein-Bar
	5	20	100.0	2721	6	E00513	E00513 Genomic DNA
	6	20	100.0	3400	6	A11128	A11128 Synthetic n
	7	20	100.0	3400	6	E01006	E01006 cDNA encodi
	8	20	100.0	3833	6	AR049357	AR049357 Sequence
	9	20	100.0	5019	6	A11178	A11178 Synthetic n
	10	20	100.0	5019	6	E01007	E01007 DNA sequenc
	11	20	100.0	5931	6	AR233080	AR233080 Sequence
	12	20	100.0	5931	14	HS4ENVGP	M10593 Epstein Bar
c	13	20	100.0	171823	14	HHV507799	AJ507799 Human her
c	14	20	100.0	172281	14	EBV	V01555 Epstein-Bar
c	15	20	100.0	184113	14	HS4B958RAJ	M80517 Epstein-Bar
c	16	17	85.0	163273	2	AC142486	AC142486 Rattus no
c	17	17	85.0	193466	2	AC113548	AC113548 Mus muscu
	18	16	80.0	204	11	G20457	G20457 human STS A
	19	16	80.0	1555	8	AY079380	AY079380 Arabidops
	20	16	80.0	1767	8	AY039920	AY039920 Arabidops
	21	16	80.0	2376	6	AX151433	AX151433 Sequence
	22	16	80.0	6203	9	AB002380	AB002380 Human mRN
	23	16	80.0	9501	9	AF180681	AF180681 Homo sapi
	24	16	80.0	16529	9	AP000271	AP000271 Homo sapi
	25	16	80.0	32813	9	AF520762	AF520762 Homo sapi
c	26	16	80.0	41308	9	AC005203	AC005203 Homo sapi
c	27	16	80.0	68589	2	AC087682	AC087682 Homo sapi
	28	16	80.0	76850	9	AC003109	AC003109 Human DNA
c	29	16	80.0	84510	6	AX646415	AX646415 Sequence
c	30	16	80.0	84510	9	AB065660	AB065660 Homo sapi
	31	16	80.0	94357	2	AC010963	AC010963 Homo sapi
c	32	16	80.0	100000	9	AP000032	AP000032 Homo sapi
	33	16	80.0	100000	9	AP000104	AP000104 Homo sapi
	34	16	80.0	100000	9	AP000180	AP000180 Homo sapi
	35	16	80.0	109149	2	AP000681	AP000681 Homo sapi
c	36	16	80.0	109741	8	AC004005	AC004005 Arabidops
	37	16	80.0	115293	9	AC127460	AC127460 Homo sapi
c	38	16	80.0	117134	9	AL157904	AL157904 Human DNA
	39	16	80.0	120836	9	AC097457	AC097457 Homo sapi
c	40	16	80.0	126626	9	AC104843	AC104843 Homo sapi
	41	16	80.0	137557	9	AC005909	AC005909 Homo sapi
	42	16	80.0	144234	2	AP002425	AP002425 Homo sapi
	43	16	80.0	144279	2	AP001570	AP001570 Homo sapi
c	44	16	80.0	151741	2	AC018609	AC018609 Mus muscu
c	45	16	80.0	153026	9	AL391280	AL391280 Human DNA
c	46	16	80.0	158621	2	AC092951	AC092951 Homo sapi
c	47	16	80.0	160133	2	AC025140	AC025140 Homo sapi
	48	16	80.0	163025	9	AP000758	AP000758 Homo sapi
c	49	16	80.0	163538	9	AC012598	AC012598 Homo sapi
c	50	16	80.0	163567	2	AL365325	AL365325 Homo sapi
	51	16	80.0	164200	9	AL591888	AL591888 Human DNA
	52	16	80.0	165106	2	AC016034	AC016034 Homo sapi
c	53	16	80.0	166305	9	AC012447	AC012447 Homo sapi
	54	16	80.0	167257	9	AC008008	AC008008 Homo sapi
c	55	16	80.0	169745	10	AL807238	AL807238 Mouse DNA
	56	16	80.0	174822	9	AC013275	AC013275 Homo sapi
c	57	16	80.0	177161	10	AC125340	AC125340 Mus muscu
c	58	16	80.0	180557	9	AC007250	AC007250 Homo sapi
	59	16	80.0	184057	10	AL928579	AL928579 Mouse DNA
	60	16	80.0	184092	9	AC073363	AC073363 Homo sapi
c	61	16	80.0	184511	9	AC009032	AC009032 Homo sapi
c	62	16	80.0	190814	9	AC006111	AC006111 Homo sapi
	63	16	80.0	192699	2	AC111102	AC111102 Mus muscu
	64	16	80.0	195485	9	AL591846	AL591846 Human DNA
c	65	16	80.0	196545	9	AC113390	AC113390 Homo sapi

66	16	80.0	197326	2	AC122380	AC122380 Mus muscu
67	16	80.0	206246	2	AL606485	AC010537 Homo sapi
68	16	80.0	214911	2	AC010537	AC010537 Homo sapi
69	16	80.0	216972	9	AC067819	AC067819 Homo sapi
70	16	80.0	217586	2	AC016076	AC016076 Homo sapi
c 71	16	80.0	229910	2	AC135698	AC135698 Rattus no
72	16	80.0	239296	2	BX323549	BX323549 Mus muscu
c 73	16	80.0	257734	2	BX510318	BX510318 Mus muscu
c 74	16	80.0	261391	2	AC094647	AC094647 Rattus no
75	16	80.0	292967	14	AF369029	AF369029 White spo
76	16	80.0	305107	6	AX151396	AX151396 Sequence
77	16	80.0	305107	14	AF332093	AF332093 White spo
78	16	80.0	307287	14	AF440570	AF440570 Shrimp wh
79	16	80.0	340000	9	AP001714	AP001714 Homo sapi
80	15	75.0	123	6	BD043816	BD043816 Sequence
c 81	15	75.0	305	1	AB106218	AB106218 Alteromon
c 82	15	75.0	654	11	BV013986	BV013986 S212P6774
83	15	75.0	702	11	BV030337	BV030337 S212P6206
84	15	75.0	713	9	HUMYZ87H02	AF086093 Homo sapi
85	15	75.0	728	11	BV056104	BV056104 S212P6629
c 86	15	75.0	762	6	AX463524	AX463524 Sequence
c 87	15	75.0	1267	3	AK117030	AK117030 Ciona int
88	15	75.0	1546	3	AB058682	AB058682 Ciona int
89	15	75.0	1619	5	AB093327	AB093327 Silurana
90	15	75.0	1648	5	AB093328	AB093328 Silurana
91	15	75.0	1691	6	AX544370	AX544370 Sequence
92	15	75.0	2171	5	DYGOP65A	M64275 Discopyge o
93	15	75.0	2556	14	AF172333	AF172333 Human her
94	15	75.0	2661	14	AF172332	AF172332 Human her
95	15	75.0	2661	14	EBVBLLF1A	X67776 Epstein-Bar
96	15	75.0	2702	1	EFPPD1GNS	X96978 E.faecalis
c 97	15	75.0	3178	5	AF034606	AF034606 Danio rer
c 98	15	75.0	3210	14	HS4GP340A	L07922 Epstein-Bar
99	15	75.0	4043	10	AF345863S2	AF345864 Mus muscu
c 100	15	75.0	4988	14	HS4GP340B	L07923 Epstein-Bar
c 101	15	75.0	5207	10	AC006119	AC006119 Mus muscu
c 102	15	75.0	8647	10	AL669922	AL669922 Mouse DNA
c 103	15	75.0	8797	10	AF028784	AF028784 Rattus no
c 104	15	75.0	16313	1	AE000879	AE000879 Methanoba
105	15	75.0	33106	10	AL928961	AL928961 Mouse DNA
c 106	15	75.0	46681	9	AC094009	AC094009 Homo sapi
c 107	15	75.0	53572	2	AC100004	AC100004 Mus muscu
c 108	15	75.0	56944	2	AC100526	AC100526 Mus muscu
c 109	15	75.0	59905	2	AC100985	AC100985 Mus muscu
110	15	75.0	64150	9	AL161778	AL161778 Human DNA
c 111	15	75.0	64739	2	AC101271	AC101271 Mus muscu
112	15	75.0	68209	9	AL357275	AL357275 Human DNA
113	15	75.0	70855	2	AC101491	AC101491 Mus muscu
114	15	75.0	71425	2	AC099894	AC099894 Mus muscu
c 115	15	75.0	72274	2	AC145201	AC145201 Takifugu
116	15	75.0	77588	2	AC018015	AC018015 Drosophil
117	15	75.0	79899	2	AC024620	AC024620 Homo sapi
c 118	15	75.0	81862	9	AL133318	AL133318 Human DNA
c 119	15	75.0	82369	10	AL929144	AL929144 Mouse DNA
c 120	15	75.0	88763	2	AC139550	AC139550 Homo sapi

ALIGNMENTS

```
RESULT 1
AX522237
LOCUS       AX522237                  20 bp    DNA     linear   PAT 24-OCT-2002
DEFINITION  Sequence 2 from Patent WO02064842.
ACCESSION   AX522237
VERSION     AX522237.1  GI:24411115
KEYWORDS    .
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1
  AUTHORS   Witte,D.P. and Groen,P.A.
  TITLE     Quantitative epstein barr virus pcr rapid assay
  JOURNAL   Patent: WO 02064842-A 2 22-AUG-2002;
            Children's Hospital Research Foundation (US)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
BASE COUNT        6 a      6 c      5 g      3 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 3.8;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db          1 CCTTAGGAGGAACAAGTCCC 20


RESULT 2
I26103
LOCUS       I26103                    20 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 29 from patent US 5556772.
ACCESSION   I26103
VERSION     I26103.1  GI:1605973
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Sorge,J.A. and Mullinax,R.L.
  TITLE     Polymerase compositions and uses thereof
  JOURNAL   Patent: US 5556772-A 29 17-SEP-1996;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        6 a      6 c      5 g      3 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 3.8;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db          1 CCTTAGGAGGAACAAGTCCC 20


RESULT 3
I04174
LOCUS       I04174                  1041 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 3 from Patent US 4707358.
ACCESSION   I04174
VERSION     I04174.1  GI:268757
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Kieff,E., Tanner,J., Hummel,M. and Beisel,C.
  TITLE     Vaccine against Epstein-Barr Virus
  JOURNAL   Patent: US 4707358-A 3 17-NOV-1987;
            The University of Chicago; Chicago, IL
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      295 a    383 c    199 g    164 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 1.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||d|||||||
Db        206 CCTTAGGAGGAACAAGTCCC 225


RESULT 4
EBVBLLF1
LOCUS       EBVBLLF1                2663 bp    DNA     linear   VRL 09-JAN-1998
DEFINITION  Epstein-Barr virus BLLF1 gene.
ACCESSION   X99106
VERSION     X99106.1  GI:2769559
KEYWORDS    BLLF1 gene; gp340.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1
  AUTHORS   Mackett,M., Cox,C., Pepper,S.D., Lees,J.F., Naylor,B.A.,
            Wedderburn,N. and Arrand,J.R.
  TITLE     Immunisation of common marmosets with vaccinia virus expressing
            Epstein-Barr virus (EBV) gp340 and challenge with EBV
  JOURNAL   J. Med. Virol. 50 (3), 263-271 (1996)
  MEDLINE   97082049
   PUBMED   8923292
REFERENCE   2  (bases 1 to 2663)
  AUTHORS   Pepper,S.D.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-1996) S.D.V. Pepper, Paterson Institute for
            Cancer Research, Molecular Biology, Wilmslow Rd, Withington,
            Manchester M20 9BX, UK
FEATURES             Location/Qualifiers
     source          1. .2663
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /strain="M81"
                     /db_xref="taxon:10376"
     gene            66. .2636
                     /gene="BLLF1"
     CDS             66. .2636
                     /gene="BLLF1"
                     /codon_start=1
                     /product="gp340"
                     /protein_id="CAA67558.1"
                     /db_xref="GI:2769560"
                     /db_xref="SPTREMBL:O56854"
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEPFPYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTQMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPRDNGTESKAPDMTS
                     PTSAVTTPTPNGTSPTPAMTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTPAVTTPT
                     PNATSPTVGETSPQANATNHTLGGTSPTPVVTSPPKNATSDVTTGQHNRTSSSTSSMS
                     LRPSSIPETTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGP
                     GNSSTSTKPGEVNVTKGTPPKNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGH
                     GARTSTEPTTDYGDDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVP
                     PTSQPRFSNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV
                     "
     polyA_signal    2641. .2646
BASE COUNT      755 a    829 c    552 g    527 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 14;  Length 2663;
  Best Local Similarity  100.0%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       1825 CCTTAGGAGGAACAAGTCCC 1844


RESULT 5
E00513
LOCUS       E00513                  2721 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Genomic DNA encoding Epstein-Barr virus gp 220/200.
ACCESSION   E00513
VERSION     E00513.1  GI:2168792
KEYWORDS    JP 1985232094-A/1.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 2721)
  AUTHORS   Eriotsuto,K., Jieroomu,T., Mearii,F. and Kurisutofuaa,B.
  TITLE     VACCINE TO EPSTAIN VAR VIRUS
  JOURNAL   Patent: JP 1985232094-A 1 18-NOV-1985;
            UNIV CHICAGO:THE
COMMENT     OS   Epstein-Barr virus
            PN   JP 1985232094-A/1
            PD   18-NOV-1985
            PF   30-JAN-1985 JP 1985014606
            PR   30-JAN-1984 US 84    575352, 23-JUL-1984 US 84    633558 PI
            ERIOTSUTO KIIFU, JIEROOMU TANNAA, MEARII FUNMERU, PI   KURISUTOFUAA
            BEISERU
            PC   C12N15/00,A61K39/245,C07H21/04,C07K13/00,C12N1/00//C12P21/02,
            PC   (C12N1/00,
            PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .>2721
            FT                   /product='gp 220/200'.
FEATURES             Location/Qualifiers
     source          1. .2721
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
BASE COUNT      762 a    876 c    557 g    526 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 2721;
  Best Local Similarity  100.0%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       1886 CCTTAGGAGGAACAAGTCCC 1905


RESULT 6
A11128
LOCUS       A11128                  3400 bp    DNA     linear   PAT 07-DEC-1993
DEFINITION  Synthetic nucleotide sequence of the leftward reading frame of the
            Bam L-fragment encoding gp 250/350.
ACCESSION   A11128
VERSION     A11128.1  GI:490970
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 3400)
  AUTHORS   Wolf,H.J.
  TITLE     DNA sequences of the EBV genome, recombinant DNA molecules,
            processes for producing EBV-related antigens, diagnostic
            compositions and pharmaceutical compositions containing said
            antigens
  JOURNAL   Patent: EP 0173254-A 3 05-MAR-1986;
            Wolf, Hans Joachim
FEATURES             Location/Qualifiers
```

```
    source          1. .3400
                    /organism="synthetic construct"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32630"
    CDS             556. .3279
                    /note="Leftward reading frame of the Bam L-fragment
                    encoding gp 250/350."
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAA00938.1"
                    /db_xref="GI:490971"
                    /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                    DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                    ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                    KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                    SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                    NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                    MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                    VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                    STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                    STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                    PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                    ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                    NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                    KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                    PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                    SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
BASE COUNT      904 a   1060 c    732 g    704 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 3400;
  Best Local Similarity  100.0%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2441 CCTTAGGAGGAACAAGTCCC 2460


RESULT 7
E01006
LOCUS       E01006                  3400 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  cDNA encoding p350 of Epstein-Barr virus.
ACCESSION   E01006
VERSION     E01006.1  GI:2169265
KEYWORDS    JP 1986257188-A/2.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 3400)
  AUTHORS   Hansu,Y.B.
  TITLE     DNA SEQUENCE OF EBV GENOM, RECOMBINED DNA MOLECULE AND EBV RELATED
            ANTIGEN AND DIAGNOSTIC COMPOSITION CONTAINING SAID ANTIGEN AND
            PRODUCTION OF PREPARATION
  JOURNAL   Patent: JP 1986257188-A 2 14-NOV-1986;
            HANSU YOTSUTO BUORUFU
COMMENT     OS   Epstein-Barr virus
            PN   JP 1986257188-A/2
            PD   14-NOV-1986
            PF   23-AUG-1985 JP 1985185661
            PR   23-AUG-1984 EP 84  84110089, 23-AUG-1984 EP 84  84110090 PI
            HANSU YOTSUTO BUORUFU
            PC   C12N15/00,A61K39/245,C07H21/04,C07K13/00,C12N1/00,G01N33/569,
            PC   G01N33/577//
            PC   C12P21/02,(C12N1/00,C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           1. .555
            FT   CDS             556. .3279
            FT                   /product='p350 Epstein-Barr virus' FT   3'UTR
                 3280. .3400
            FT   polyA_site      3284. .3291.
FEATURES            Location/Qualifiers
     source         1. .3400
                    /organism="Human herpesvirus 4"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10376"
BASE COUNT      904 a   1060 c    732 g    704 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 3400;
  Best Local Similarity  100.0%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2441 CCTTAGGAGGAACAAGTCCC 2460


RESULT 8
AR049357
LOCUS       AR049357                3833 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 18 from patent US 5824508.
ACCESSION   AR049357
VERSION     AR049357.1  GI:6005396
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3833)
  AUTHORS   Spaete,R. and Jackman,W.T.
  TITLE     Non-splicing variants of gp350/220
  JOURNAL   Patent: US 5824508-A 18 20-OCT-1998;
FEATURES            Location/Qualifiers
     source         1. .3833
                    /organism="unknown"
BASE COUNT     1013 a   1165 c    846 g    809 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 3833;
  Best Local Similarity  100.0%;  Pred. No. 1.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 9
A11178
LOCUS       A11178                  5019 bp    DNA     linear   PAT 07-DEC-1993
DEFINITION  Synthetic nucleotide sequence for the fusion protein encoded by
            pURLP1.9.
ACCESSION   A11178
VERSION     A11178.1  GI:490997
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 5019)
  AUTHORS   Wolf,H.J.
  TITLE     DNA sequences of the EBV genome, recombinant DNA molecules,
            processes for producing EBV-related antigens, diagnostic
            compositions and pharmaceutical compositions containing said
            antigens
  JOURNAL   Patent: EP 0173254-A 5 05-MAR-1986;
            Wolf, Hans Joachim
FEATURES            Location/Qualifiers
     source         1. .5019
                    /organism="synthetic construct"
                    /mol_type="genomic DNA"
```

```
                /db_xref="taxon:32630"
     CDS        4. .5019
                /note="Fusion protein encoded by pURLP1.9."
                /codon_start=1
                /transl_table=11
                /protein_id="CAA00945.1"
                /db_xref="GI:490998"
                /translation="MITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTD
                RPSQQLRSLNGEWRFAWFPAPEAVPESWLECDLPEADTVVVPSNWQMHGYDAPIYTNV
                TYPITVNPPFVPTENPTGCYSLTFNVDESWLQEGQTRIIFDGVNSAFHLWCNGRWVGY
                GQDSRLPSEFDLSAFLRAGENRLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKPT
                TQISDFHVATRFNDDFSRAVLEAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGE
                IIDERGGYADRVTLRLNVENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFREV
                RIENGLLLLNGKPLLIRGVNRHEHHPLHGQVMDEQTMVQDILLMKQNNFNAVRCSHYP
                NHPLWYTLCDRYGLYVVDEANIETHGMVPMNRLTDDPRWLPAMSERVTRMVQRDRNHP
                SVIIWSLGNESGHGANHDALYRWIKSVDPSRPVQYEGGGADTTATDIICPMYARVDED
                QPFPAVPKWSIKKWLSLPGETRPLILCEYAHAMGNSLGGFAKYWQAFRQYPRLQGGFV
                WDWVDQSLIKYDENGNPWSAYGGDFGDTPNDRQFCMNGLVFADRTPHPALTEAKHQQQ
                FFQFRLSGQTIEVTSEYLFRHSHNELLHWMVALDGKPLASGEVPLDVAPQGKQLIELP
                ELPQPESAGQLWLTVRVVQPNATAWSEAGHISAWQQWRLAENLSVTLPAASHAIPHLT
                TSEMDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTPLRDQFTRAPLDNDIGVSEAT
                RIDPNAWVERWKAAGHYQAEAALLQCTADTLADAVLITTAHAWQHQGKTLFISRKTYR
                IDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERVNWLGLGPQENYPDRLTAACF
                DRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQWRGDFQFNISRYSQQQLMETSHR
                HLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVWCQKGDPSTCSG
                YESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLGNNSILYVFYSGNGPKA
                SGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVPMVTSEDANSPNVTVTA
                FWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDITVSGLGTAPKTLIITRT
                ATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPASTGP
                TVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPT
                PAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSA
                VTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQANATNHTLGGTSPTPVVT
                SQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTG
                GENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNA
                TSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRY
                NATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNLSMLVLQPSLSMIS
                CQT"
BASE COUNT     1230 a   1530 c   1265 g    994 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 5019;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       4277 CCTTAGGAGGAACAAGTCCC 4296


RESULT 10
E01007
LOCUS       E01007                  5019 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence of hybrid peptide which contains beta-gelactosidase
            and p350 of Epstein-Barr virus.
ACCESSION   E01007
VERSION     E01007.1  GI:2169266
KEYWORDS    JP 1986257188-A/3.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 5019)
  AUTHORS   Hansu,Y.B.
  TITLE     DNA SEQUENCE OF EBV GENOM, RECOMBINED DNA MOLECULE AND EBV RELATED
            ANTIGEN AND DIAGNOSTIC COMPOSITION CONTAINING SAID ANTIGEN AND
            PRODUCTION OF PREPARATION
  JOURNAL   Patent: JP 1986257188-A 3 14-NOV-1986;
            HANSU YOTSUTO BUORUFU
COMMENT     OS   Artificial gene
            OC   Artificial sequence; Genes.
            PN   JP 1986257188-A/3
            PD   14-NOV-1986
            PF   23-AUG-1985 JP 1985185661
            PR   23-AUG-1984 EP 84  84110089, 23-AUG-1984 EP 84  84110090 PI
            HANSU YOTSUTO BUORUFU
            PC   C12N15/00,A61K39/245,C07H21/04,C07K13/00,C12N1/00,G01N33/569,
            PC   G01N33/577//
            PC   C12P21/02,(C12N1/00,C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: clone=pURLP1.9;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           1. .3
            FT   CDS             4. .3069
            FT                   /product='beta-galactosidase' FT
            misc_feature    3070. .3072
            FT                   /note='pUR290 linker'
            FT   misc_feature    3073. .3088
            FT                   /note='poly cloning region'
            FT   CDS             3089. .4985
            FT                   /product='gp350 of Epstein-Barr virus' FT
            misc_feature    4986. .4994
            FT                   /note='polycloning region'
            FT   misc_feature    4995. .5019
            FT                   /note='pBR322'.
FEATURES             Location/Qualifiers
     source          1. .5019
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT     1230 a   1529 c   1266 g    994 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 5019;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       4277 CCTTAGGAGGAACAAGTCCC 4296


RESULT 11
AR233080
LOCUS       AR233080                5931 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6458364.
ACCESSION   AR233080
VERSION     AR233080.1  GI:27275511
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5931)
  AUTHORS   Spaete,R. and Jackman,W.T.
  TITLE     Non-splicing variants of gp350/220
  JOURNAL   Patent: US 6458364-A 1 01-OCT-2002;
FEATURES             Location/Qualifiers
     source          1. .5931
                     /organism="unknown"
BASE COUNT     1451 a   1788 c   1433 g   1259 t
ORIGIN

  Query Match           100.0%;  Score 20;  DB 6;  Length 5931;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 12
HS4ENVGP
LOCUS       HS4ENVGP                5931 bp    DNA     linear   VRL 02-AUG-1993
```

```
DEFINITION  Epstein Barr virus major outer envelope glycoprotein genes gp350
            and gp220, complete cds.
ACCESSION   M10593
VERSION     M10593.1  GI:330360
KEYWORDS    envelope protein; glycoprotein; glycoprotein 220; glycoprotein 350.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 5931)
  AUTHORS   Beisel,C., Tanner,J., Matsuo,T., Thorley-Lawson,D., Kezdy,F. and
            Kieff,E.
  TITLE     Two major outer envelope glycoproteins of Epstein-Barr virus are
            encoded by the same gene
  JOURNAL   J. Virol. 54 (3), 665-674 (1985)
  MEDLINE   85211017
   PUBMED   2987520
COMMENT     Original source text: Epstein Barr virus, DNA, clones III57 and
            pKH47 (passed in B95-8 cells).
            Clean copy sequence for [1] kindly provided by E.Kieff,
            16-DEC-1985.
            The two major outer envelope glycoproteins, gp350 and gp220 start
            and end in the same places, but gp220 has an intron that gp350 does
            not.
            A potential TATA box is located at positions 975-981 and a
            potential poly-A signal at 3742-3747.
FEATURES             Location/Qualifiers
     source          1. .5931
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
     mRNA            995. .3771
                     /product="gp350, gp220 mRNA"
     CDS             1014. .3737
                     /note="major outer envelope glycoprotein gp350"
                     /codon_start=1
                     /protein_id="AAA45881.1"
                     /db_xref="GI:330362"
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                     STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                     PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                     ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                     NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPEPRPGTTSQASGPGNSSTST
                     KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                     PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                     SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
     CDS             join(1014. .2514,3262. .3737)
                     /note="major outer envelope glycoprotein gp220"
                     /codon_start=1
                     /protein_id="AAA45880.1"
                     /db_xref="GI:330361"
                     /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                     DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                     ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                     KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                     SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                     NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                     MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                     VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                     STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESTPPQNAT
                     SPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYN
                     ATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRFSNLSMLVLQWASLAVLT
                     LLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
     exon            <1014. .2514
                     /note="major outer envelope glycoprotein gp220"
     intron          2515. .3261
                     /note="gp220 cds intron"
     exon            3262. .>3737
                     /note="major outer envelope glycoprotein gp220"
                     /number=2
BASE COUNT    1453 a   1784 c   1437 g   1257 t
ORIGIN      1 bp upstream of EcoRI site.

  Query Match           100.0%;  Score 20;  DB 14;  Length 5931;
  Best Local Similarity  100.0%;  Pred. No. 1.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db       2899 CCTTAGGAGGAACAAGTCCC 2918


RESULT 13
HHV507799/c
LOCUS       HHV507799             171823 bp    DNA     circular VRL 04-APR-2003
DEFINITION  Human herpesvirus 4 complete wild type genome.
ACCESSION   AJ507799
VERSION     AJ507799.1  GI:23893576
KEYWORDS    complete genome.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1
  AUTHORS   Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and
            Griffin,B.E.
  TITLE     Molecular cloning of the complete Epstein-Barr virus genome as a
            set of overlapping restriction endonuclease fragments
  JOURNAL   Nucleic Acids Res. 9 (13), 2999-3014 (1981)
  MEDLINE   82014887
   PUBMED   6269068
REFERENCE   2
  AUTHORS   Kozak,M.
  TITLE     Possible role of flanking nucleotides in recognition of the AUG
            initiator codon by eukaryotic ribosomes
  JOURNAL   Nucleic Acids Res. 9 (20), 5233-5262 (1981)
  MEDLINE   82059504
   PUBMED   7301588
REFERENCE   3
  AUTHORS   Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
  TITLE     Sequence analysis and in vitro transcription of portions of the
            Epstein-Barr virus genome
  JOURNAL   J. Cell. Biochem. 19 (3), 267-274 (1982)
  MEDLINE   83109311
   PUBMED   6296170
REFERENCE   4
  AUTHORS   Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE     Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
  MEDLINE   83169725
   PUBMED   6300857
REFERENCE   5
  AUTHORS   Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE     Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
            Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (1), 21-45 (1983)
  MEDLINE   85035713
   PUBMED   6092825
REFERENCE   6
  AUTHORS   Seguin,C., Farrell,P.J. and Barrell,B.G.
  TITLE     DNA sequence and transcription of the BamHI fragment B region of
            B95-8 Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (3), 369-392 (1983)
  MEDLINE   85060424
   PUBMED   6094953
REFERENCE   7
  AUTHORS   Jeang,K.T. and Hayward,S.D.
  TITLE     Organization of the Epstein-Barr virus DNA molecule. III. Location
```

of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
JOURNAL J. Virol. 48 (1), 135-148 (1983)
MEDLINE 83294686
PUBMED 6310141
REFERENCE 8
AUTHORS Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.
TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences
JOURNAL Mol. Biol. Med. 1 (4), 425-445 (1983)
MEDLINE 85060428
PUBMED 6094955
REFERENCE 9
AUTHORS Farrell,P.J., Bankier,A., Seguin,C., Deininger,P. and Barrell,B.G.
TITLE Latent and lytic cycle promoters of Epstein-Barr virus
JOURNAL EMBO J. 2 (8), 1331-1338 (1983)
MEDLINE 20331131
PUBMED 10872327
REFERENCE 10
AUTHORS Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
JOURNAL EMBO J. 3 (4), 813-821 (1984)
MEDLINE 84207939
PUBMED 6327290
REFERENCE 11
AUTHORS Biggin,M., Farrell,P.J. and Barrell,B.G.
TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus
JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
MEDLINE 84236104
PUBMED 6203743
REFERENCE 12
AUTHORS Yates,J., Warren,N., Reisman,D. and Sugden,B.
TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
MEDLINE 84222045
PUBMED 6328526
REFERENCE 13
AUTHORS Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
MEDLINE 84247360
PUBMED 6330697
REFERENCE 14 (bases 1 to 171823)
AUTHORS Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G.
TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
JOURNAL Nature 310 (5974), 207-211 (1984)
MEDLINE 84270667
PUBMED 6087149
REFERENCE 15
AUTHORS Bodescot,M. and Perricaudet,M.
TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
MEDLINE 87289053
PUBMED 3039467
REFERENCE 16
AUTHORS Laux,G., Perricaudet,M. and Farrell,P.J.
TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
JOURNAL EMBO J. 7 (3), 769-774 (1988)
MEDLINE 88283646
PUBMED 2840285
REFERENCE 17
AUTHORS Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
TITLE Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region
JOURNAL Virology 179 (1), 339-346 (1990)
MEDLINE 91021036
PUBMED 2171209
REFERENCE 18 (bases 1 to 171823)
AUTHORS Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
JOURNAL Unpublished
REFERENCE 19
AUTHORS Binne,U.K., Amon,W. and Farrell,P.J.
TITLE Induction of Epstein-Barr virus late promoters on small plasmids in the EBV late lytic cycle requires ori lyt
JOURNAL Unpublished
REFERENCE 20 (bases 1 to 171823)
AUTHORS Farrell,P.J.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2002) Farrell P., Ludwig Institute for Cancer Research, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place London W2 1PG
COMMENT Construction:
This sequence was assembled from B95-8 EBV [14] and Raji EBV [18] with sequence corrections [16, 19]. The number of major internal repeat units has been reduced from 11.6 [14] to a more typical 7.6 and the B95-8 deletion sequences have been restored to give a sequence more representative of wild type EBV.
Numbering
Like the modified B95-8 sequence[14, 16] accession number V01555, this sequence starts 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC.).

FEATURES Location/Qualifiers

source 1. .171823
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/strain="B95-8"
/db_xref="taxon:10376"

source join(1. .139223,151555. .171823)
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/strain="B95-8"
/db_xref="taxon:10376"

source 139224. .151554
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/strain="Raji"
/db_xref="taxon:10376"

gene join(166103. .171823,1. .1682)
/gene="LMP2"

CDS join(166103. .166458,58. .272,360. .458,540. .788,871. .951, 1026. .1196,1280. .1495,1574. .1680)
/gene="LMP2"
/codon_start=1
/product="terminal protein LMP2A"
/protein_id="CAD53382.1"
/db_xref="GI:23893577"
/translation="MGSLEMVPMGAGPPSPGGDPDGYDGGNNSQYPSASGSSGNTPTP
PNDEERESNEEPPPPYEDPYWGNGDRHSDYQPLGTQDQSLYLGLQHDGNDGLPPPPYS
PRDDSSQHIYEEAGRGSMNPVCLPVIVAPYLFWLAAIAASCFTASVSTVVTATGLALS
LLLLAAVASSYAAAQRKLLTPVTVLTAVVTFFAICLTWRIEDPPFNSLLFALLAAAGG
LQGIYVLVMLVLLILAYRRRWRRLTVCGGIMFLACVLVLIVDAVLQLSPLLGAVTVVS
MTLLLLAFVLWLSSPGGLGTLGAALLTLAAALALLASLILGTLNLTTMFLLMLLWTLV
VLLICSSCSSCPLSKILLARLFLYALALLLLASALIAGGSILQTNFKSLSSTEFIPNL
FCMLLLIVAGILFILAILTEWGSGNRTYGPVFMCLGGLLTMVAGAVWLTVMSNTLLSA
WILTAGFLIFLIGFALFGVIRCCRYCCYYCLTLESEERPPTPYRNTV"

exon 58. .272
/gene="LMP2"
/number=2

CDS join(59. .272,360. .458,540. .788,871. .951,1026. .1196, 1280. .1495,1574. .1680)
/gene="LMP2"
/codon_start=1
/product="terminal protein LMP2B"
/protein_id="CAD53383.1"

```
  Query Match             100.0%;  Score 20;  DB 14;  Length 171823;
  Best Local Similarity   100.0%;  Pred. No. 0.55;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db      77980 CCTTAGGAGGAACAAGTCCC 77961


RESULT 14
EBV/c
LOCUS       EBV                     172281 bp    DNA     circular VRL 20-SEP-1999
DEFINITION  Epstein-Barr virus (EBV) genome, strain B95-8.
ACCESSION   V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784
VERSION     V01555.1  GI:59074
KEYWORDS    DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem
            repeat; terminal repeat.
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 172281)
  AUTHORS   Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and
            Griffin,B.E.
  TITLE     Molecular cloning of the complete Epstein-Barr virus genome as a
            set of overlapping restriction endonuclease fragments
  JOURNAL   Nucleic Acids Res. 9 (13), 2999-3014 (1981)
  MEDLINE   82014887
   PUBMED   6269068
REFERENCE   2  (bases 1 to 172281)
  AUTHORS   Kozak,M.
  TITLE     Possible role of flanking nucleotides in recognition of the AUG
            initiator codon by eukaryotic ribosomes
  JOURNAL   Nucleic Acids Res. 9 (20), 5233-5262 (1981)
  MEDLINE   82059504
   PUBMED   7301588
REFERENCE   3  (bases 1 to 172281)
  AUTHORS   Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
  TITLE     Sequence analysis and in vitro transcription of portions of the
            Epstein-Barr virus genome
  JOURNAL   J. Cell. Biochem. 19 (3), 267-274 (1982)
  MEDLINE   83109311
   PUBMED   6296170
REFERENCE   4  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE     Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
  MEDLINE   83169725
   PUBMED   6300857
REFERENCE   5  (bases 142687 to 159853)
  AUTHORS   Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE     Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
            Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (1), 21-45 (1983)
  MEDLINE   85035713
   PUBMED   6092825
REFERENCE   6  (bases 112620 to 125316)
  AUTHORS   Seguin,C., Farrell,P.J. and Barrell,B.G.
  TITLE     DNA sequence and transcription of the BamHI fragment B region of
            B95-8 Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (3), 369-392 (1983)
  MEDLINE   85060424
   PUBMED   6094953
REFERENCE   7  (bases 45644 to 52450)
  AUTHORS   Jeang,K.T. and Hayward,S.D.
  TITLE     Organization of the Epstein-Barr virus DNA molecule. III. Location
            of the P3HR-1 deletion junction and characterization of the NotI
            repeat units that form part of the template for an abundant
            12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
  JOURNAL   J. Virol. 48 (1), 135-148 (1983)
  MEDLINE   83294686
   PUBMED   6310141
REFERENCE   8  (bases 159853 to 172281)
  AUTHORS   Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J.
            and Barrell,B.G.
  TITLE     DNA sequence analysis of the EcoRI Dhet fragment of B95-8
            Epstein-Barr virus containing the terminal repeat sequences
  JOURNAL   Mol. Biol. Med. 1 (4), 425-445 (1983)
  MEDLINE   85060428
   PUBMED   6094955
REFERENCE   9  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Bankier,A., Seguin,C., Deininger,P. and Barrell,B.G.
  TITLE     Latent and lytic cycle promoters of Epstein-Barr virus
  JOURNAL   EMBO J. 2 (8), 1331-1338 (1983)
  MEDLINE   20331131
   PUBMED   10872327
REFERENCE   10 (bases 45415 to 52824)
  AUTHORS   Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
  TITLE     The EB virus genome in Daudi Burkitt's lymphoma cells has a
            deletion similar to that observed in a non-transforming strain
            (P3HR-1) of the virus
  JOURNAL   EMBO J. 3 (4), 813-821 (1984)
  MEDLINE   84207939
   PUBMED   6327290
REFERENCE   11 (bases 87650 to 92703)
  AUTHORS   Biggin,M., Farrell,P.J. and Barrell,B.G.
  TITLE     Transcription and DNA sequence of the BamHI L fragment of B95-8
            Epstein-Barr virus
  JOURNAL   EMBO J. 3 (5), 1083-1090 (1984)
  MEDLINE   84236104
   PUBMED   6203743
REFERENCE   12 (bases 7315 to 9312)
  AUTHORS   Yates,J., Warren,N., Reisman,D. and Sugden,B.
  TITLE     A cis-acting element from the Epstein-Barr viral genome that
            permits stable replication of recombinant plasmids in latently
            infected cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
  MEDLINE   84222045
   PUBMED   6328526
REFERENCE   13 (bases 76089 to 79808)
  AUTHORS   Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
  TITLE     Homology between two EBV early genes and HSV ribonucleotide
            reductase and 38K genes
  JOURNAL   Nucleic Acids Res. 12 (12), 5087-5099 (1984)
  MEDLINE   84247360
   PUBMED   6330697
REFERENCE   14 (bases 1 to 172281)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310 (5974), 207-211 (1984)
  MEDLINE   84270667
   PUBMED   6087149
REFERENCE   15 (bases 1 to 172281)
  AUTHORS   Bodescot,M. and Perricaudet,M.
  TITLE     Clustered alternative splice sites in Epstein-Barr virus RNAs
  JOURNAL   Nucleic Acids Res. 15 (14), 5887 (1987)
  MEDLINE   87289053
   PUBMED   3039467
REFERENCE   16 (bases 1 to 172281)
  AUTHORS   Laux,G., Perricaudet,M. and Farrell,P.J.
  TITLE     A spliced Epstein-Barr virus gene expressed in immortalized
            lymphocytes is created by circularization of the linear viral
            genome
  JOURNAL   EMBO J. 7 (3), 769-774 (1988)
  MEDLINE   88283646
   PUBMED   2840285
REFERENCE   17 (bases 1 to 172281)
  AUTHORS   Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
  JOURNAL   Unpublished
REFERENCE   18 (bases 1 to 172281)
  AUTHORS   Farrell,P.J. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1984)
```

```
REFERENCE   19 (bases 1 to 172281)
  AUTHORS   Farrell,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer
            Research, St. Mary's Hospital Medical School, Norfolk Place London
            W2 1PG
COMMENT     CDS
            Listed under this feature are all known protein coding regions as
            well as all the major open reading frames in the sequence. In
            general the term major is taken as the longest frame in a
            particular region taking into account the adjacent longest frames
            and likely transcription signals. Note that on this basis some long
            overlapping frames have been excluded and on the other hand some
            small frames have been included which might represent exons or
            genes because they occur in a logical combination with other
            features or because of some other experimental data. The reading
            frames are named according to the Bam H1 fragment in which they
            start. eg BALF3 is the third leftward frame starting in Bam H1
            fragment A. BORF1 is the first rightward frame in Bam H1 fragment
            O. If there is an obvious TATA sequence followed by an in frame Met
            codon that satisfies the rules of Kozak [12] in that there is a
            purine at -3 and/or a G at +4 then the reading frame is numbered
            from the A of the ATG to the base preceding the termination codon.
            If there is no obvious initiation codon or there is a substantial
            reading frame in phase before the ATG then the reading frame is
            numbered from the first base of the first codon.
            SITEs of POLYA signals
            This feature lists all occurences of the sequence AATAAA which is
            found normally approximately 20 bases upstream of the mRNA
            processing/polyA addition site. The rarely used homolog ATTAAA is
            only listed when it is found in a position close to the end of a
            major reading frame.
            SITEs of DONOR and ACCEPT sequences
            This is not a comprehensive listing of all such sequences and only
            the positions of a few have been noted because they occur in
            potentially interesting positions. The number quoted in the table
            is the position of the terminal base in the intron in each case.
            Restriction enzyme SITEs.
            Only the positions of the sites Bam HI (BAM) are listed.
            RPT
            This feature is used to define repetitive sequences.
            SITE DEL
            This feature defines deletions in B95-8 with respect to other
            strains such as RAJI and also to deletions in other strains such as
            P3HR1 and DAUDI with respect to B95-8.
            SITE HPN
            Denotes sequences with twofold symmetry ie could form hairpin
            loops. This is not a comprehensive list - only a few occurences
            noted.
            ORGRPL
            Denotes the region that encompasses an origin of replication (ori
            P).[13].
            NUMBERING
            The DNA sequence of B95-8 EBV has been revised [19]. The original
            (Baer et al, 1984) base 359 has been deleted so the new sequence
            around that position reads TCAGTCTTT. To avoid renumbering the
            entire sequence, position 1 has benn moved 1 base to the left of
            the EcoRI site separating EcoRI Dhet from EcoRI I
            (ie the first A of AGAATTC).
FEATURES             Location/Qualifiers
     source          1. .172281
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /strain="B95-8"
                     /db_xref="taxon:10376"
     mRNA            58. .272

  Query Match            100.0%;  Score 20;  DB 14;  Length 172281;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db      90268 CCTTAGGAGGAACAAGTCCC 90249

RESULT 15
HS4B958RAJ/c
LOCUS       HS4B958RAJ            184113 bp    DNA     linear   VRL 12-APR-1996
DEFINITION  Epstein-Barr virus, artifactual joining of B95-8 complete genome
            and the sequences from Raji of the large deletion found in B95-8.
ACCESSION   M80517 M75989
VERSION     M80517.1  GI:330330
KEYWORDS    .
SOURCE      Human herpesvirus 4 (Epstein-Barr virus)
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (sites)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310 (5974), 207-211 (1984)
  MEDLINE   84270667
   PUBMED   6087149
REFERENCE   2  (sites)
  AUTHORS   Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
  TITLE     Sequence and transcription of Raji Epstein-Barr virus DNA spanning
            the B95-8 deletion region
  JOURNAL   Virology 179 (1), 339-346 (1990)
  MEDLINE   91021036
   PUBMED   2171209
REFERENCE   3  (sites)
  AUTHORS   Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,
            Rickinson,A. and Kieff,E.
  TITLE     Restricted Epstein-Barr virus protein expression in Burkitt
            lymphoma is due to a different Epstein-Barr nuclear antigen 1
            transcriptional initiation site
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)
  MEDLINE   91296817
   PUBMED   1648738
REFERENCE   4  (bases 1 to 184113)
  AUTHORS   Jenson,H.B.
  TITLE     GenBank Curator Program
  JOURNAL   Unpublished (1992)
COMMENT     Original source text: Human herpesvirus 4 DNA.
            The B95-8 genome (V01555) has a large deletion in the right side of
            the genome which has been sequenced in Raji (M35547).  These
            sequences have been joined to form an extended and more complete,
            although artifactual, EBV sequence.
            For features, refer to feature tables of V01555 and M35547.
FEATURES             Location/Qualifiers
     source          1. .184113
                     /organism="Human herpesvirus 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10376"
     misc_feature    1. .152008
                     /note="B95-8 sequences (corresponds to 1-152,008 of
                     V01555)"
     misc_feature    152009. .152012
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 in V01555,
                     and 1-4 in M35547)"
     misc_feature    153013. .163839
                     /note="Raji sequences (corresponds to 5-11,831 of M35547)"
     misc_feature    163840. .163843
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 of V01555,
                     and 11,832-11,835 of M35547)"
     misc_feature    163844. .184113
                     /note="B95-8 sequences (corresponds to 152,013-172,282 of
                     V01555)"
BASE COUNT    36002 a  55824 c  54622 g  37665 t
ORIGIN
```

```
  Query Match           100.0%;  Score 20;  DB 14;  Length 184113;
  Best Local Similarity  100.0%;  Pred. No. 0.54;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGTCCC 20
              ||||||||||||||||||||
Db      90268 CCTTAGGAGGAACAAGTCCC 90249


RESULT 16
AC142486/c
LOCUS       AC142486               163273 bp    DNA     linear   HTG 01-APR-2003
DEFINITION  Rattus norvegicus clone CH230-522E23, *** SEQUENCING IN PROGRESS
            ***, 49 unordered pieces.
ACCESSION   AC142486
VERSION     AC142486.2  GI:29423830
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 163273)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163273)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163273)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 1, 2003 this sequence version replaced gi:29374196.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KEMQ
                Center clone name: CH230-522E23
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 140865 bases at least Q40
                Consensus quality: 147625 bases at least Q30
                Consensus quality: 152515 bases at least Q20
                Estimated insert size: 149404; sum-of-contigs estimation
                Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 49 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1102: contig of 1102 bp in length
            *     1103     1202: gap of unknown length
            *     1203     2433: contig of 1231 bp in length
            *     2434     2533: gap of unknown length
            *     2534     3667: contig of 1134 bp in length
            *     3668     3767: gap of unknown length
            *     3768     4781: contig of 1014 bp in length
            *     4782     4881: gap of unknown length
            *     4882     6332: contig of 1451 bp in length
            *     6333     6432: gap of unknown length
            *     6433     7726: contig of 1294 bp in length
            *     7727     7826: gap of unknown length
            *     7827     9239: contig of 1413 bp in length
            *     9240     9339: gap of unknown length
            *     9340    10646: contig of 1307 bp in length
            *    10647    10746: gap of unknown length
            *    10747    13007: contig of 2261 bp in length
            *    13008    13107: gap of unknown length
            *    13108    14650: contig of 1543 bp in length
            *    14651    14750: gap of unknown length
            *    14751    16039: contig of 1289 bp in length
            *    16040    16139: gap of unknown length
            *    16140    18232: contig of 2093 bp in length
            *    18233    18332: gap of unknown length
            *    18333    19866: contig of 1534 bp in length
            *    19867    19966: gap of unknown length
            *    19967    21078: contig of 1112 bp in length
            *    21079    21178: gap of unknown length
            *    21179    22965: contig of 1787 bp in length
            *    22966    23065: gap of unknown length
            *    23066    25141: contig of 2076 bp in length
            *    25142    25241: gap of unknown length
```

```
          *    25242    26668: contig of 1427 bp in length
          *    26669    26768: gap of unknown length
          *    26769    28506: contig of 1738 bp in length
          *    28507    28606: gap of unknown length
          *    28607    30429: contig of 1823 bp in length
          *    30430    30529: gap of unknown length
          *    30530    32570: contig of 2041 bp in length
          *    32571    32670: gap of unknown length
          *    32671    34029: contig of 1359 bp in length
          *    34030    34129: gap of unknown length
          *    34130    36411: contig of 2282 bp in length
          *    36412    36511: gap of unknown length
          *    36512    38475: contig of 1964 bp in length
          *    38476    38575: gap of unknown length
          *    38576    40562: contig of 1987 bp in length
          *    40563    40662: gap of unknown length
          *    40663    42221: contig of 1559 bp in length
          *    42222    42321: gap of unknown length
          *    42322    45005: contig of 2684 bp in length
          *    45006    45105: gap of unknown length
          *    45106    46300: contig of 1195 bp in length
          *    46301    46400: gap of unknown length
          *    46401    48411: contig of 2011 bp in length
          *    48412    48511: gap of unknown length
          *    48512    51544: contig of 3033 bp in length
          *    51545    51644: gap of unknown length
          *    51645    55142: contig of 3498 bp in length
          *    55143    55242: gap of unknown length
          *    55243    58168: contig of 2926 bp in length
          *    58169    58268: gap of unknown length
          *    58269    60879: contig of 2611 bp in length
          *    60880    60979: gap of unknown length
          *    60980    63070: contig of 2091 bp in length
          *    63071    63170: gap of unknown length
          *    63171    65129: contig of 1959 bp in length
          *    65130    65229: gap of unknown length
          *    65230    68519: contig of 3290 bp in length
          *    68520    68619: gap of unknown length
          *    68620    71820: contig of 3201 bp in length
          *    71821    71920: gap of unknown length
          *    71921    77299: contig of 5379 bp in length
          *    77300    77399: gap of unknown length
          *    77400    81762: contig of 4363 bp in length
          *    81763    81862: gap of unknown length
          *    81863    86382: contig of 4520 bp in length
          *    86383    86482: gap of unknown length
          *    86483    90542: contig of 4060 bp in length
          *    90543    90642: gap of unknown length
          *    90643    95533: contig of 4891 bp in length
          *    95534    95633: gap of unknown length
          *    95634   101752: contig of 6119 bp in length
          *   101753   101852: gap of unknown length
          *   101853   108908: contig of 7056 bp in length
          *   108909   109008: gap of unknown length
          *   109009   115481: contig of 6473 bp in length
          *   115482   115581: gap of unknown length
          *   115582   122306: contig of 6725 bp in length
          *   122307   122406: gap of unknown length
          *   122407   130001: contig of 7595 bp in length
          *   130002   130101: gap of unknown length
          *   130102   138191: contig of 8090 bp in length
          *   138192   138291: gap of unknown length
          *   138292   150672: contig of 12381 bp in length
          *   150673   150772: gap of unknown length
          *   150773   163273: contig of 12501 bp in length.

  Query Match             85.0%;  Score 17;  DB 2;  Length 163273;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGTC 18
              |||||||||||||||||
Db      66325 CTTAGGAGGAACAAGTC 66309
```

```
RESULT 17
AC113548/c
LOCUS       AC113548              193466 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Mus musculus clone RP23-268F15, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC113548
VERSION     AC113548.5  GI:28913170
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 193466)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-268F15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193466)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 193466)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2003 this sequence version replaced gi:28626751.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
```

```
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          ------------- Project Information
            Center project name: L24242
            Center clone name: 268_F_15
          ------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 191143 bases at least Q40
            Consensus quality: 191818 bases at least Q30
            Consensus quality: 192162 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 192466; sum-of-contigs
            Quality coverage: 8.5 in Q20 bases; agarose-fp
            Quality coverage: 8.5 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     4933: contig of 4933 bp in length
          *     4934     5033: gap of 100 bp
          *     5034     6097: contig of 1064 bp in length
          *     6098     6197: gap of 100 bp
          *     6198     7378: contig of 1181 bp in length
          *     7379     7478: gap of 100 bp
          *     7479     9778: contig of 2300 bp in length
          *     9779     9878: gap of 100 bp
          *     9879    11949: contig of 2071 bp in length
          *    11950    12049: gap of 100 bp
          *    12050    15641: contig of 3592 bp in length
          *    15642    15741: gap of 100 bp
          *    15742    19456: contig of 3715 bp in length
          *    19457    19556: gap of 100 bp
          *    19557    39905: contig of 20349 bp in length
          *    39906    40005: gap of 100 bp
          *    40006    67129: contig of 27124 bp in length
          *    67130    67229: gap of 100 bp
          *    67230   104769: contig of 37540 bp in length
          *   104770   104869: gap of 100 bp
          *   104870   193466: contig of 88597 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193466
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-268F15"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .4933
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5034. .6097
                     /note="assembly_fragment"
     misc_feature    6198. .7378
                     /note="assembly_fragment"
     misc_feature    7479. .9778
                     /note="assembly_fragment"
     misc_feature    9879. .11949
                     /note="assembly_fragment"
     misc_feature    12050. .15641
                     /note="assembly_fragment"
     misc_feature    15742. .19456
                     /note="assembly_fragment"
     misc_feature    19557. .39905
                     /note="assembly_fragment"
     misc_feature    40006. .67129
                     /note="assembly_fragment"
     misc_feature    67230. .104769
                     /note="assembly_fragment"
     misc_feature    104870. .193466
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53362 a  45171 c  42934 g  50990 t   1009 others
ORIGIN

  Query Match             85.0%;  Score 17;  DB 2;  Length 193466;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TAGGAGGAACAAGTCCC 20
              |||||||||||||||||
Db     147268 TAGGAGGAACAAGTCCC 147252


RESULT 18
G20457
LOCUS       G20457                    204 bp    DNA     linear   STS 24-JUL-1996
DEFINITION  human STS A005R48, sequence tagged site.
ACCESSION   G20457
VERSION     G20457.1  GI:1340794
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204)
  AUTHORS   Adams,M.D.
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Mark Adams
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Email: mdadams@tigr.org

            Primer A: GTATCTCTTAGGAGGAACAAGTA
            Primer B: ATTGAGTGCTAGAGAATTACATA
            STS size: 204
            PCR Profile:
                      Denaturation:  96C 5min
                      Anneal:        54C 30sec
                      Extend:        72C 30sec
                      Denature:      95C 30sec
                      FinalExtend:   72C  5min
                      Cycles:        30
            Protocol:
                    GenomicDNA:   25 ng
                    Primer:        0.43 uM each
                    dNTPs:       230 uM each
                    AmpliTaq:      0.5 units
                    TaqStart Ab:   0.5 units
                    Total Volume: 10 ul

            Buffer:
                      Tris-HCl pH8.8: 100 mM
                      KCl:            500 mM
                      MgCl2:           20 mM
                      Triton X-100:    1%
                      Concentration:  10X

            Prepared with primer pairs derived from THC105423: GenBank
            Accession Numbers-  H80345, H70110, F11453, F13120, R24326, R24340,
            R24852, T75305, Z42603.
FEATURES             Location/Qualifiers
     source          1. .204
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
     STS             1. .204
     primer_bind     1. .23
     primer_bind     complement(182. .204)
BASE COUNT       76 a     39 c     34 g     55 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 11;  Length 204;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db          7 CTTAGGAGGAACAAGT 22


RESULT 19
AY079380
LOCUS       AY079380               1555 bp    mRNA    linear   PLN 18-SEP-2002
DEFINITION  Arabidopsis thaliana unknown protein (At2g44020) mRNA, complete
            cds.
ACCESSION   AY079380
VERSION     AY079380.1  GI:19310760
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1555)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Etgu,P.,
            Lee,J.M., Toriumi,M., Yu,G., Brooks,S., Chao,Q., Chen,H.,
            Karlin-Neumann,G., Kim,C., Lam,B., Miranda,M., Nguyen,M.,
            Palm,C.J., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Arabidopsis Open Reading Frame (ORF) Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1555)
  AUTHORS   Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
            Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     The RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
            Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
            Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
            Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H.,
            Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
            Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
            to this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
            /PGEC) contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES             Location/Qualifiers
     source          1. .1555
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="U09491"
                     /note="This clone is in pUNI 51.
                     ecotype: Columbia"
     gene            1. .1555
                     /gene="At2g44020"
     CDS             1. .1524
                     /gene="At2g44020"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAL85111.1"
                     /db_xref="GI:19310761"
                     /translation="MSYLLRRNKFVALLKRQSLIFPITSTEAKTLINPDPNIPQFQNP
                     CSIFRIAHYATQSSKFPEYEMPTVTWGVIQGKKEKLVNRVKICDYLKGLGIITDELES
                     IELPSTIEVMCERVEFLQKLGLTIDDINEYPLMLGCSVRKNLIPVLAYLEKIGISRSK
                     LGEFVKNYPQVLHASVVVELAPVVKFLRGLDVEKQDLGYVLMKYPELLGFKLEGTMST
                     SVAYLVSIGVSPRDIGPMVTQYPYLLGMRVGTMIKPLVDYLISIGLPKKIVARMLEKR
                     SYIVGYNLEETVKPNVDCLISFGVKKELLPLLIAQYPQILGLPVKAKMSTQQYFFSLK
                     LKIDPEGFARVVEKMPQIVSLKQNVIMKPIEFLLGRAFQVEDIAKMVVRCPQILCSRV
                     ELMKNSYYFYKTEMGRPMKELVEYPEYFTYSLESRIKPRYQKLQSKGIRSSLNWFLNC
                     SDQRFEERLQGNFIDPDTEGPTFDMGGKLEMPGGEIVTDEEEDESDDEVLYRRTLTL"
     3'UTR           1525. .1555
                     /gene="At2g44020"
BASE COUNT       463 a    271 c    372 g    449 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 8;  Length 1555;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         13 CTTAGGAGGAACAAGT 28


RESULT 20
AY039920
LOCUS       AY039920               1767 bp    mRNA    linear   PLN 18-SEP-2002
DEFINITION  Arabidopsis thaliana unknown protein (At2g44020) mRNA, complete
            cds.
ACCESSION   AY039920
VERSION     AY039920.1  GI:14532591
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1767)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1767)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Dale,J.M., Gibson,H.A., Goldsmith,A.D., Jiang,P.X., Lee,J.M.,
            Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L., Carninci,P.,
            Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
```

```
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
            Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Dale,J.M., Gibson,H.A.,
            Goldsmith,A.D., Jiang,P.X., Lee,J.M., Quach,H.L., Tang,C.C.,
            Toriumi,M., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T.,
            Karlin-Neumann,G., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES                Location/Qualifiers
     source             1. .1767
                        /organism="Arabidopsis thaliana"
                        /mol_type="mRNA"
                        /db_xref="taxon:3702"
                        /chromosome="2"
                        /clone="RAFL04-13-P03 (R09491)"
                        /note="This clone is in a modified pBluescript vector
                        (Lambda ZAP) as a XhoI/SstI insert.
                        ecotype: Columbia"
     gene               1. .1767
                        /gene="At2g44020"
     5'UTR              1. .125
                        /gene="At2g44020"
     CDS                126. .1649
                        /gene="At2g44020"
                        /codon_start=1
                        /evidence=experimental
                        /product="unknown protein"
                        /protein_id="AAK64024.1"
                        /db_xref="GI:14532592"
                        /translation="MSYLLRRNKFVALLKRQSLIFPITSTEAKTLINPDPNIPQFQNP
                        CSIFRIAHYATQSSKFPEYEMPTVTWGVIQGKKEKLVNRVKICDYLKGLGIITDELES
                        IELPSTIEVMCERVEFLQKLGLTIDDINEYPLMLGCSVRKNLIPVLAYLEKIGISRSK
                        LGEFVKNYPQVLHASVVVELAPVVKFLRGLDVEKQDLGYVLMKYPELLGFKLEGTMST
                        SVAYLVSIGVSPRDIGPMVTQYPYLLGMRVGTMIKPLVDYLISIGLPKKIVARMLEKR
                        SYIVGYNLEETVKPNVDCLISFGVKKELLPLLIAQYPQILGLPVKAKMSTQQYFFSLK
                        LKIDPEGFARVVEKMPQIVSLKQNVIMKPIEFLLGRAFQVEDIAKMVVRCPQILCSRV
                        ELMKNSYYFYKTEMGRPMKELVEYPEYFTYSLESRIKPRYQKLQSKGIRSSLNWFLNC
                        SDQRFEERLQGNFIDPDTEGPTFDMGGKLEMPGGEIVTDEEEDESDDEVLYRRTLTL"
     3'UTR              1650. .1767
                        /gene="At2g44020"
BASE COUNT      520 a    316 c    404 g    527 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 8;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        138 CTTAGGAGGAACAAGT 153


RESULT 21
AX151433
LOCUS       AX151433               2376 bp    DNA     linear   PAT 22-JUN-2001
DEFINITION  Sequence 38 from Patent WO0138351.
ACCESSION   AX151433
VERSION     AX151433.1  GI:14533498
KEYWORDS    .
SOURCE      Shrimp white spot syndrome virus
  ORGANISM  Shrimp white spot syndrome virus
            Viruses; dsDNA viruses, no RNA stage; Nimaviridae.
REFERENCE   1
  AUTHORS   Xu,X., Yang,F., He,J., Pham,L.Z., He,M., Ye,Y., Shen,Y. and
            Kodira,C.
  TITLE     Nucleotide sequence of the shrimp white spot syndrome bacilliform
            virus (wsbv), systems containing this sequence and detection kits
  JOURNAL   Patent: WO 0138351-A 38 31-MAY-2001;
            PE Corporation (NY) (US) ; The Third Institute of Oceanography,
            State Oceanic Administration (CN) ; Sinogenomax Co., Ltd. (CN)
FEATURES                Location/Qualifiers
     source             1. .2376
                        /organism="Shrimp white spot syndrome virus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:92652"
BASE COUNT      787 a    467 c    526 g    596 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 6;  Length 2376;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        793 AGGAGGAACAAGTCCC 808


RESULT 22
AB002380
LOCUS       AB002380               6203 bp    mRNA    linear   PRI 06-OCT-2001
DEFINITION  Human mRNA for KIAA0382 gene, partial cds.
ACCESSION   AB002380
VERSION     AB002380.1  GI:2224704
KEYWORDS    KIAA0382.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
   PUBMED   9205841
REFERENCE   2  (bases 1 to 6203)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
FEATURES                Location/Qualifiers
     source             1. .6203
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="HH0685"
                        /sex="male"
                        /tissue_type="brain"
                        /clone_lib="pBluescriptII SK plus"
     gene               1. .6203
                        /gene="KIAA0382"
     CDS                <1. .2255
                        /gene="KIAA0382"
                        /codon_start=3
                        /protein_id="BAA20836.1"
                        /db_xref="GI:2224705"
                        /translation="LFYTERAHVRTLKVLDQVFYQRVSREGILSPSELRKIFSNLEDI
                        LQLHIGLNEQMKAVRKRNETSVIDQIGEDLLTWFSGPGEEKLKHAAATFCSNQPFALE
                        MIKSRQKKDSRFQTFVQDAESNPLCRRLQLKDIIPTQMQRLTKYPLLLDNIAKYTEWP
                        TEREKVKKAADHCRQILNFVNQAVKEAENKQRLEDYQRRLDTSSLKLSEYPNVEELRN
                        LDLTKRKMIHEGPLVWKVNRDKTIDLYTLLLEDILVLLQKQDDRLVLRCHSKILASTA
```

DSKHTFSPVIKLSTVLVRQVATDNKALFVISMSDNGAQIYELVAQTVSEKTVWQDLIC
RMAASVKEQSTKPIPLPQSTPGEGDNDEEDPSKLKEEQHGISVTGLQSPDRDLGLEST
LISSKPQSHSLSTSGKSEVRDLFVAERQFAKEQHTDGTLKEVGEDYQIAIPDSHLPVS
EERWALDALRNLGLLKQLLVQQLGLTEKSVQEDWQHFPRYRTASQGPQTDSVIQNSEN
IKAYHSGEGHMPFRTGTGDIATCYSPRTSTESFAPRDSVGLAPQDSQASNILVMDHMI
MTPEMPTMEPEGGLDDSGEHFFDAREAHSDENPSEGDGAVNKEEKDVNLRISGNYLIL
DGYDPVQESSTDEEVASSLTLQPMTGIPAVESTHQQQHSPQNTHSDGAISPFTPEFLV
QQRWGAMEYSCFEIQSPSSCADSQSQIMEYIHKIEADLEHLKKVEESYTILCQRLAGS
ALTDKHSDKS"

BASE COUNT 1808 a 1195 c 1336 g 1864 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 6203;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 4168 CTTAGGAGGAACAAGT 4183

RESULT 23
AF180681
LOCUS AF180681 9501 bp mRNA linear PRI 02-MAR-2000
DEFINITION Homo sapiens guanine nucleotide exchange factor (LARG) mRNA, complete cds.
ACCESSION AF180681
VERSION AF180681.1 GI:7110159
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9501)
AUTHORS Kourlas,P.J., Strout,M.P., Becknell,B., Veronese,M.L., Croce,C.M., Theil,K.S., Krahe,R., Ruutu,T., Knuutila,S., Bloomfield,C.D. and Caligiuri,M.A.
TITLE Identification of a gene at 11q23 encoding a guanine nucleotide exchange factor: evidence for its fusion with MLL in acute myeloid leukemia
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (5), 2145-2150 (2000)
MEDLINE 20160919
PUBMED 10681437
REFERENCE 2 (bases 1 to 9501)
AUTHORS Kourlas,P.J., Strout,M.P., Becknell,B., Bloomfield,C.B. and Caligiuri,M.A.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1999) Internal Medicine, Division of Hematology-Oncology, The Ohio State University, 4th floor Starling-Loving Hall, 320 West 10th Ave, Columbus, OH 43210, USA
FEATURES Location/Qualifiers
source 1. .9501
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23.3"
/tissue_type="prostate"
gene 1. .9501
/gene="LARG"
CDS 8. .4642
/gene="LARG"
/codon_start=1
/product="guanine nucleotide exchange factor"
/protein_id="AAF36817.1"
/db_xref="GI:7110160"
/translation="MSGTQSTITDRFPLKKPIRHGSILNRESPTDKKQKVERIASHDF
DPTDSSSKKTKSSSEESRSEIYGLVQRCVIIQKDDNGFGLTVSGDNPVFVQSVKEDGA
AMRAGVQTGDRIIKVNGTLVTHSNHLEVVKLIKSGSYVALTVQGRPPGSPQIPLADSE
VEPSVIGHMSPIMTSPHSPGASGNMERITSPVLMGEENNVVHNQKVEILRKMLQKEQE
RLQLLQEDYNRTPAQRLLKEIQEAKKHIPQLQEQLSKATGSAQDGAVVTPSRPLGDTL
TVSEAETDPGDVLGRTDCSSGDASRPSSDNADSPKSGPKERIYLEENPEKSETIQDTD
TQSLVGSPSTRIAPHIIGAEDDDFGTEHEQINGQCSCFQSIELLKSRPAHLAVFLHHV
VSQFDPATLLCYLYSDLYKHTNSKETRRIFLEFHQFFLDRSAHLKVSVPDEMSADLEK
RRPELIPEDLHRHYIQTMQERVHPEVQRHLEDFRQKRSMGLTLAESELTKLDAERDKD
RLTLEKERTCAEQIVAKIEEVLMTAQAVEEDKSSTMQYVILMYMKHLGVKVKEPRNLE
HKRGRIGFLPKIKQSMKKDKEGEEKGKRRGFPSILGPPRRPSRHDNSAIGRAMELQKA
RHPKHLSTPSSVSPEPQDSAKLRQSGLANEGTDAGYLPANSMSSVASGASFSQEGGKE
NDTGSKQVGETSAPGDTLDGTPRTLNTVFDFPPPPLDQVQEEECEVERVTEHGTPKPF
RKFDSVAFGESQSEDEQFENDLETDPPNWQQLVSREVLLGLKPCEIKRQEVINELFYT
ERAHVRTLKVLDQVFYQRVSREGILSPSELRKIFSNLEDILQLHIGLNEQMKAVRKRN
ETSVIDQIGEDLLTWFSGPGEEKLKHAAATFCSNQPFALEMIKSRQKKDSRFQTFVQD
AESNPLCRRLQLKDIIPTQMQRLTKYPLLLDNIAKYTEWPTEREKVKKAADHCRQILN
YVNQAVKEAENKQRLEDYQRRLDTSSLKLSEYPNVEELRNLDLTKRKMIHEGPLVWKV
NRDKTIDLYTLLLEDILVLLQKQDDRLVLRCHSKILASTADSKHTFSPVIKLSTVLVR
QVATDNKALFVISMSDNGAQIYELVAQTVSEKTVWQDLICRMAASVKEQSTKPIPLPQ
STPGEGDNDEEDPSKLKEEQHGISVTGLQSPDRDLGLESTLISSKPQSHSLSTSGKSE
VRDLFVAERQFAKEQHTDGTLKEVGEDYQIAIPDSHLPVSEERWALDALRNLGLLKQL
LVQQLGLTEKSVQEDWQHFPRYRTASQGPQTDSVIQNSENIKAYHSGEGHMPFRTGTG
DIATCYSPRTSTESFAPRDSVGLAPQDSQASNILVMDHMIMTPEMPTMEPEGGLDDSG
EHFFDAREAHSDENPSEGDGAVNKEEKDVNLRISGNYLILDGYDPVQESSTDEEVASS
LTLQPMTGIPAVESTHQQQHSPQNTHSDGAISPFTPEFLVQQRWGAMEYSCFEIQSPS
SCADSQSQIMEYIHKIEADLEHLKKVEESYTILCQRLAGSALTDKHSDKS"

BASE COUNT 2816 a 1872 c 2078 g 2735 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 9501;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 6556 CTTAGGAGGAACAAGT 6571

RESULT 24
AP000271
LOCUS AP000271 16529 bp DNA linear PRI 12-JUN-2003
DEFINITION Homo sapiens genomic DNA, chromosome 21 clone:CMF21-74A4, complete sequence.
ACCESSION AP000271
VERSION AP000271.3 GI:31621048
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 16529)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Jun 12, 2003 this sequence version replaced gi:31071640.
FEATURES Location/Qualifiers
source 1. .16529
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q"
/clone="CMF21-74A4"

BASE COUNT 4172 a 3739 c 3784 g 4834 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 16529;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db       5368 AGGAGGAACAAGTCCC 5383


RESULT 25
AF520762
LOCUS       AF520762               32813 bp    DNA     linear   PRI 20-JUN-2002
DEFINITION  Homo sapiens X-ray repair complementing defective repair in Chinese
            hamster cells 2 (XRCC2) gene, complete cds.
ACCESSION   AF520762
VERSION     AF520762.1  GI:21489901
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32813)
  AUTHORS   Rieder,M.J., Livingston,R.J., Braun,A.C., Montoya,M.A.,
            Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L.,
            Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and
            Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2002) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use:  NIEHS-SNPs, Environmental Genome
            Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
            (URL: http://egp.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .32813
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   23. .335
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       77
                     /frequency="0.03"
                     /replace="a"
     variation       297
                     /frequency="0.05"
                     /replace="a"
     variation       332
                     /frequency="0.18"
                     /replace="c"
     variation       353
                     /frequency="0.05"
                     /replace="c"
     variation       353
                     /frequency="0.36"
                     /replace="a"
     repeat_region   432. .635
                     /rpt_family="MaLR"
                     /rpt_type=dispersed
     variation       523
                     /frequency="0.01"
                     /replace="t"
     variation       583
                     /frequency="0.18"
                     /replace="t"
     variation       612
                     /frequency="0.36"
                     /replace="g"
     variation       683
                     /frequency="0.01"
                     /replace="c"
     repeat_region   809. .1057
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       899
                     /frequency="0.01"
                     /replace="a"
     variation       924
                     /frequency="0.04"
                     /replace="a"
     variation       964
                     /frequency="0.05"
                     /replace="g"
     variation       977
                     /frequency="0.05"
                     /replace="c"
     misc_feature    1030. .1130
                     /note="Region not scanned for variation"
     variation       1133. .1136
                     /frequency="0.91"
                     /replace=""
     variation       1171
                     /frequency="0.01"
                     /replace="t"
     variation       1334
                     /frequency="0.04"
                     /replace="t"
     variation       1389
                     /frequency="0.01"
                     /replace="c"
     variation       1498
                     /frequency="0.18"
                     /replace="c"
     gene            1500. .31217
                     /gene="XRCC2"
     mRNA            join(1500. .1624,16885. .16966,28340. .31217)
                     /gene="XRCC2"
                     /product="X-ray repair complementing defective repair in
                     Chinese hamster cells 2"
     variation       1517
                     /gene="XRCC2"
                     /frequency="0.06"
                     /replace="g"
     CDS             join(1586. .1624,16885. .16966,28340. .29061)
                     /gene="XRCC2"
                     /codon_start=1
                     /product="X-ray repair complementing defective repair in
                     Chinese hamster cells 2"
                     /protein_id="AAM55241.1"
                     /db_xref="GI:21489902"
                     /translation="MCSAFHRAESGTELLARLEGRSSLKEIEPNLFADEDSPVHGDIL
                     EFHGPEGTGKTEMLYHLTARCILPKSEGGLEVEVLFIDTDYHFDMLRLVTILEHRLSQ
                     SSEEIIKYCLGRFFLVYCSSSTHLLLTLYSLESMFCSHPSLCLLILDLSLAFYWIDRV
                     NGGESVNLQESTLRKCSQCLEKLVNDYRLVLFATTQTIMQKASSSSEEPSHASRRLCD
                     VDIDYRPYLCKAWQQLVKHRMFFSKQDDSQSSNQFSLVSRCLKSNSLKKHFFIIGESG
                     VEFC"
     variation       1826
                     /gene="XRCC2"
                     /frequency="0.44"
                     /replace="t"
     variation       1929
                     /gene="XRCC2"
                     /frequency="0.02"
                     /replace="a"
     variation       1934
                     /gene="XRCC2"
                     /frequency="0.05"
                     /replace="c"
     variation       2002
                     /gene="XRCC2"
                     /frequency="0.03"
                     /replace="g"
     variation       2150
                     /gene="XRCC2"
                     /frequency="0.12"
                     /replace="c"
     variation       2177
                     /gene="XRCC2"
                     /frequency="0.02"
                     /replace="a"
```

```
     misc_feature    2556. .3762
                     /gene="XRCC2"
                     /note="Region not scanned for variation"
     repeat_region   2618. .2814
                     /rpt_family="MER1_type"
                     /rpt_type=dispersed
     repeat_region   3047. .3331
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   3774. .4079
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       3810
                     /gene="XRCC2"
                     /frequency="0.11"
                     /replace="t"
     variation       3864
                     /gene="XRCC2"
                     /frequency="0.01"
                     /replace="t"
     variation       3902
                     /gene="XRCC2"
                     /frequency="0.01"
                     /replace="t"
     variation       3982
                     /gene="XRCC2"
                     /frequency="0.03"
                     /replace="g"
     repeat_region   4201. .4476
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       4285
                     /gene="XRCC2"
                     /frequency="0.04"
                     /replace="a"
     variation       4522
                     /gene="XRCC2"
                     /frequency="0.04"
                     /replace="t"
     variation       4554
                     /gene="XRCC2"
                     /frequency="0.04"
                     /replace="t"
     variation       4558
                     /gene="XRCC2"
                     /frequency="0.05"
                     /replace="t"
     variation       4576
                     /gene="XRCC2"
                     /frequency="0.01"
                     /replace="c"
     variation       4896
                     /gene="XRCC2"
                     /frequency="0.01"
                     /replace="a"
     repeat_region   5038. .5209
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     variation       5062
                     /gene="XRCC2"
                     /frequency="0.05"
                     /replace="a"

  Query Match             80.0%;  Score 16;  DB 9;  Length 32813;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      24451 AGGAGGAACAAGTCCC 24466


RESULT 26
AC005203/c
LOCUS       AC005203               41308 bp    DNA     linear   PRI 30-JUN-1998
DEFINITION  Homo sapiens chromosome 16, cosmid clone RT70 (LANL), complete
            sequence.
ACCESSION   AC005203
VERSION     AC005203.1  GI:3273381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 41308)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41308)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Sequencing of Human Chromosome 16p13.3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 41308)
  AUTHORS   Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
            Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
            Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
            White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
            Misra,M. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1998) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
FEATURES             Location/Qualifiers
     source          1. .41308
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.3"
                     /clone="RT70"
     misc_feature    complement(4. .321)
                     /note="97% identity EST zd49c11.s1"
                     /db_xref="dbEST:W70228"
     STS             complement(4. .321)
                     /standard_name="SHGC-57870"
                     /note="97% identity"
                     /db_xref="dbSTS:G37491"
     repeat_region   44. .213
                     /rpt_family="MLT1"
     repeat_region   272. .732
                     /rpt_family="Alu"
     repeat_region   654. .686
                     /note="(A)33"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   717. .981
                     /rpt_family="Alu"
     prim_transcript 1017. .1142
                     /standard_name="fmfc19"
                     /note="93% identity"
     prim_transcript complement(1433. .1698)
                     /standard_name="fmfc18"
                     /note="99% identity"
     gene            1747. .4885
                     /gene="DNaseI"
                     /note="DNaseI"
     exon            <1747. .1889
                     /gene="DNaseI"
     exon            2218. .2365
                     /gene="DNaseI"
     CDS             join(2219. .2365,2694. .2782,2947. .3030,3539. .3654,
```

```
                     3900. .4012,4088. .4242,4621. .4717,4807. .4854)
                     /gene="DNaseI"
                     /codon_start=1
                     /evidence=experimental
                     /product="DNaseI"
                     /protein_id="AAC24721.1"
                     /db_xref="GI:3273382"
                     /translation="MRGMKLLGALLALAALLQGAVSLKIAAFNIQTFGETKMSNATLV
                     SYIVQILSRYDIALVQEVRDSHLTAVGKLLDNLNQDAPDTYHYVVSEPLGRNSYKERY
                     LFVYRPDQVSAVDSYYYDDGCEPCGNDTFNREPAIVRFFSRFTEVREFAIVPLHAAPG
                     DAVAEIDALYDVYLDVQEKWGLEDVMLMGDFNAGCSYVRPSQWSSIRLWTSPTFQWLI
                     PDSADTTATPTHCAYDRIVVAGMLLRGAVVPDSALPFNFQAAYGLSDQLAQAISDHYP
                     VEVMLK"
exon                 2694. .2782
                     /gene="DNaseI"
exon                 2947. .3030
                     /gene="DNaseI"
exon                 3539. .3654
                     /gene="DNaseI"
exon                 3900. .4012
                     /gene="DNaseI"
exon                 4088. .4242
                     /gene="DNaseI"
exon                 4621. .4717
                     /gene="DNaseI"
exon                 4807. .>4885
                     /gene="DNaseI"
STS                  4939. .5135
                     /standard_name="WI-6217"
                     /db_xref="dbSTS:G06033"
gene                 complement(4945. .>37887)
                     /gene="TRAP1"
                     /note="tumor necrosis factor type 1 receptor associated
                     protein"
exon                 complement(4945. .5131)
                     /gene="TRAP1"
STS                  4968. .5135
                     /standard_name="TIGR-A002F42"
                     /db_xref="dbSTS:G26214"
repeat_region        complement(5177. .5342)
                     /rpt_family="Alu"
repeat_region        5347. .5600
                     /rpt_family="Alu"
exon                 complement(5694. .5766)
                     /gene="TRAP1"
repeat_region        complement(6633. .6929)
                     /rpt_family="Alu"
repeat_region        6651. .6682
                     /note="(TAGA)8"
                     /rpt_type=tandem
                     /rpt_unit=TAGA
repeat_region        7223. .7360
                     /rpt_family="Alu"
repeat_region        7643. .7846
                     /rpt_family="Alu"
prim_transcript      complement(join(8887. .9035,9772. .9863))
                     /gene="TRAP1"
                     /standard_name="mj70b08"
                     /note="88% & 83% identity mouse dbEST:AA060574"
CDS                  complement(join(<8888. .9033,9783. .9868,10325. .10463,
                     11175. .11360,12872. .13019,18624. .18693,19602. .19722,
                     21241. .21396,22226. .22299,22938. .23047,24400. .24560,
                     26621. .26692,32898. .33038,35957. .36039,37729. .>37887))
                     /gene="TRAP1"
                     /note="tumor necrosis factor type 1 receptor associated
                     protein"
                     /codon_start=3
                     /evidence=experimental
                     /product="TRAP1"
                     /protein_id="AAC24722.1"
                     /db_xref="GI:3273383"
                     /translation="KPILCPRRTTAQLGPRRNPAWSLQAGRLFSTQTAEDKEEPLHSI
                     ISSTESVQGSTSKHEFQAETKKLLDIVARSLYSEKEVFIRELISNASDALEKLRHKLV
                     SDGQALPEMEIHLQTNAEKGTITIQDTGIGMTQEELVSNLGTIARSGSKAFLDALQNQ
                     AEASSKIIGQFGVGFYSAFMVADRVEVYSRSAAPGSLGYQWLSDGSGVFEIAEASGVR
                     TGTKIIIHLKSDCKEFSSEARVRDVVTKYSNFVSFPLYLNGRRMNTLQAIWMMDPKDV
                     REWQHEEFYRYVAQAHDKPRYTLHYKTDAPLNIRSIFYVPDMKPSMFDVSRELGSSVA
                     LYSRKVLIQTKATDILPKWLRFIRGVVDSEDIPLNLSRELLQESALIRKLRDVLQQRL
                     IKFFIDQSKKDAEKYAKFFEDYGLFMREGIVTATEQEVKEDIAKLLRYESSALPSGQL
                     TSLSEYASRMRAGTRNIYYLCAPNRHLAEHSPYYEAMKKKDTEVLFCFEQFDELTLLH
                     LREFDKKKLISVETDIVVDHYKEEKFEDRSPAAECLSEKETEELMAWMRNVLGSRVTN
                     VKVTLRLDTHPAMVTVLEMGAARHFLRMQQLAKTQEERAQLLQPTLEINP"
exon                 complement(8888. .9033)
                     /gene="TRAP1"
prim_transcript      complement(join(8891. .9035,9772. .9876))
                     /gene="TRAP1"
                     /standard_name="ve64f11"
                     /note="87% & 80% identity mouse EST dbEST:AA438165"
exon                 complement(9783. .9868)
                     /gene="TRAP1"
exon                 complement(10325. .10463)
                     /gene="TRAP1"
repeat_region        10693. .10995
                     /rpt_family="Alu"
exon                 complement(11175. .11360)
                     /gene="TRAP1"
exon                 complement(12872. .13019)
                     /gene="TRAP1"
repeat_region        14231. .14572
                     /rpt_family="Alu"
repeat_region        14528. .14551
                     /note="(AAAT)6"
                     /rpt_type=tandem
                     /rpt_unit=AAAT
repeat_region        complement(14642. .15261)
                     /rpt_family="Alu"
repeat_region        14828. .14849
                     /note="(T)22"
                     /rpt_type=tandem
                     /rpt_unit=T
repeat_region        16449. .16764
                     /rpt_family="Alu"
repeat_region        16735. .16762
                     /note="(A)28"
                     /rpt_type=tandem
                     /rpt_unit=A
repeat_region        16990. .17632
                     /rpt_family="Alu"
repeat_region        17285. .17313
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
exon                 complement(18624. .18693)
                     /gene="TRAP1"
repeat_region        18932. .19247
                     /rpt_family="Alu"
exon                 complement(19602. .19722)
                     /gene="TRAP1"

  Query Match             80.0%;  Score 16;  DB 9;  Length 41308;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      34206 AGGAGGAACAAGTCCC 34191


RESULT 27
AC087682/c
LOCUS       AC087682               68589 bp    DNA     linear   HTG 15-JAN-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-886K22 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC087682
VERSION     AC087682.1  GI:12229457
KEYWORDS    HTG; HTGS_PHASE0.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68589)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-886K22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68589)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12306
                Center clone name: 886_K_22
            --------------.
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      714: contig of 714 bp in length
            *      715      814: gap of 100 bp
            *      815     1522: contig of 708 bp in length
            *     1523     1622: gap of 100 bp
            *     1623     2330: contig of 708 bp in length
            *     2331     2430: gap of 100 bp
            *     2431     3132: contig of 702 bp in length
            *     3133     3232: gap of 100 bp
            *     3233     3944: contig of 712 bp in length
            *     3945     4044: gap of 100 bp
            *     4045     4741: contig of 697 bp in length
            *     4742     4841: gap of 100 bp
            *     4842     5570: contig of 729 bp in length
            *     5571     5670: gap of 100 bp
            *     5671     6387: contig of 717 bp in length
            *     6388     6487: gap of 100 bp
            *     6488     7189: contig of 702 bp in length
            *     7190     7289: gap of 100 bp
            *     7290     7992: contig of 703 bp in length
            *     7993     8092: gap of 100 bp
            *     8093     8796: contig of 704 bp in length
            *     8797     8896: gap of 100 bp
            *     8897     9603: contig of 707 bp in length
            *     9604     9703: gap of 100 bp
            *     9704    10405: contig of 702 bp in length
            *    10406    10505: gap of 100 bp
            *    10506    11206: contig of 701 bp in length
            *    11207    11306: gap of 100 bp
            *    11307    12042: contig of 736 bp in length
            *    12043    12142: gap of 100 bp
            *    12143    12855: contig of 713 bp in length
            *    12856    12955: gap of 100 bp
            *    12956    13654: contig of 699 bp in length
            *    13655    13754: gap of 100 bp
            *    13755    14489: contig of 735 bp in length
            *    14490    14589: gap of 100 bp
            *    14590    15269: contig of 680 bp in length
            *    15270    15369: gap of 100 bp
            *    15370    16083: contig of 714 bp in length
            *    16084    16183: gap of 100 bp
            *    16184    16887: contig of 704 bp in length
            *    16888    16987: gap of 100 bp
            *    16988    17677: contig of 690 bp in length
            *    17678    17777: gap of 100 bp
            *    17778    18483: contig of 706 bp in length
            *    18484    18583: gap of 100 bp
            *    18584    19299: contig of 716 bp in length
            *    19300    19399: gap of 100 bp
            *    19400    20119: contig of 720 bp in length
            *    20120    20219: gap of 100 bp
            *    20220    20945: contig of 726 bp in length
            *    20946    21045: gap of 100 bp
            *    21046    21763: contig of 718 bp in length
            *    21764    21863: gap of 100 bp
            *    21864    22556: contig of 693 bp in length
            *    22557    22656: gap of 100 bp
            *    22657    23375: contig of 719 bp in length
            *    23376    23475: gap of 100 bp
            *    23476    24179: contig of 704 bp in length
            *    24180    24279: gap of 100 bp
            *    24280    24973: contig of 694 bp in length
            *    24974    25073: gap of 100 bp
            *    25074    25788: contig of 715 bp in length
            *    25789    25888: gap of 100 bp
            *    25889    26602: contig of 714 bp in length
            *    26603    26702: gap of 100 bp
            *    26703    27398: contig of 696 bp in length
            *    27399    27498: gap of 100 bp
            *    27499    28190: contig of 692 bp in length
            *    28191    28290: gap of 100 bp
            *    28291    29008: contig of 718 bp in length
            *    29009    29108: gap of 100 bp
            *    29109    29825: contig of 717 bp in length
            *    29826    29925: gap of 100 bp
            *    29926    30628: contig of 703 bp in length
            *    30629    30728: gap of 100 bp
            *    30729    31426: contig of 698 bp in length
            *    31427    31526: gap of 100 bp
            *    31527    32222: contig of 696 bp in length
            *    32223    32322: gap of 100 bp
            *    32323    33015: contig of 693 bp in length
            *    33016    33115: gap of 100 bp
            *    33116    33832: contig of 717 bp in length
            *    33833    33932: gap of 100 bp
            *    33933    34659: contig of 727 bp in length
            *    34660    34759: gap of 100 bp
            *    34760    35450: contig of 691 bp in length
            *    35451    35550: gap of 100 bp
            *    35551    36270: contig of 720 bp in length
            *    36271    36370: gap of 100 bp
```

```
            *    36371    37078: contig of 708 bp in length
            *    37079    37178: gap of 100 bp
            *    37179    37892: contig of 714 bp in length
            *    37893    37992: gap of 100 bp
            *    37993    38715: contig of 723 bp in length
            *    38716    38815: gap of 100 bp
            *    38816    39528: contig of 713 bp in length
            *    39529    39628: gap of 100 bp
            *    39629    40340: contig of 712 bp in length
            *    40341    40440: gap of 100 bp
            *    40441    41084: contig of 644 bp in length
            *    41085    41184: gap of 100 bp
            *    41185    41877: contig of 693 bp in length
            *    41878    41977: gap of 100 bp
            *    41978    42655: contig of 678 bp in length
            *    42656    42755: gap of 100 bp
            *    42756    43464: contig of 709 bp in length
            *    43465    43564: gap of 100 bp
            *    43565    44301: contig of 737 bp in length
            *    44302    44401: gap of 100 bp
            *    44402    45116: contig of 715 bp in length
            *    45117    45216: gap of 100 bp
            *    45217    45927: contig of 711 bp in length
            *    45928    46027: gap of 100 bp
            *    46028    46702: contig of 675 bp in length
            *    46703    46802: gap of 100 bp
            *    46803    47515: contig of 713 bp in length
            *    47516    47615: gap of 100 bp
            *    47616    48326: contig of 711 bp in length
            *    48327    48426: gap of 100 bp
            *    48427    49125: contig of 699 bp in length
            *    49126    49225: gap of 100 bp
            *    49226    49930: contig of 705 bp in length
            *    49931    50030: gap of 100 bp
            *    50031    50744: contig of 714 bp in length
            *    50745    50844: gap of 100 bp
            *    50845    51534: contig of 690 bp in length
            *    51535    51634: gap of 100 bp
            *    51635    52348: contig of 714 bp in length
            *    52349    52448: gap of 100 bp
            *    52449    53192: contig of 744 bp in length
            *    53193    53292: gap of 100 bp
            *    53293    53930: contig of 638 bp in length
            *    53931    54030: gap of 100 bp
            *    54031    54736: contig of 706 bp in length
            *    54737    54836: gap of 100 bp
            *    54837    55531: contig of 695 bp in length
            *    55532    55631: gap of 100 bp

  Query Match              80.0%;  Score 16;  DB 2;  Length 68589;
  Best Local Similarity    100.0%;  Pred. No. 88;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      63479 AGGAGGAACAAGTCCC 63464


RESULT 28
AC003109
LOCUS       AC003109               76850 bp    DNA     linear   PRI 17-MAR-1998
DEFINITION  Human DNA from overlapping chromosome 7 PAC and P1 clones
            containing the XRCC2 gene, genomic sequence, complete sequence.
ACCESSION   AC003109
VERSION     AC003109.1  GI:2961444
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76850)
  AUTHORS   Liu,N., Lamerdin,J.E., Tebbs,R.S., Schild,D., Tucker,J.D., Shen,R.,
            Brookman,K.W., Siciliano,M.J., Walter,C.A., Fan,W., Narayana,L.S.,
            Zhou,Z.-Q., Adamson,A.W., Sorenson,K.J., Chen,D.J., Jones,N.J. and
            Thompson,L.H.
  TITLE     XRCC2 and XRCC3, New Human Rad51- Family Members, Promote
            Chromosome Stability and Protect Against DNA Crosslinks and Other
            Damages
  JOURNAL   Mol. Cell (1998) In press
REFERENCE   2  (bases 1 to 76850)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 76850)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Sequence derived from PAC clone 13620 from bases 1 to 10,382, and
            from P1 clone 7515 from bases 10,095 to 76,850. No polymorphisms or
            discrepancies were detected in the overlapping 288 base pairs.
FEATURES             Location/Qualifiers
     source          1. .76850
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q36"
                     /clone="hsXRCC2"
                     /note="P1 clone 7515 obtained from Genome Systems (library
                     constructed from male foreskin-derived fibroblast line).
                     PAC clone 13620 obtained from Genome Systems (library
                     constructed from male leukocytes by P. de Jong)."
     repeat_region   2. .208
                     /rpt_family="Alu"
     repeat_region   complement(209. .272)
                     /rpt_family="MER42"
     repeat_region   complement(380. .442)
                     /rpt_family="Alu"
     repeat_region   complement(497. .813)
                     /rpt_family="L1"
     repeat_region   complement(825. .1106)
                     /rpt_family="Alu"
     repeat_region   1178. .1464
                     /rpt_family="Alu"
     repeat_region   complement(1789. .1899)
                     /rpt_family="L1"
     repeat_region   2034. .2157
                     /rpt_family="Alu"
     repeat_region   2731. .3063
                     /rpt_family="Alu"
     repeat_region   complement(3549. .3789)
                     /rpt_family="Alu"
     mRNA            4236. .33903
                     /product="XRCC2"
                     /note="3'-UTR contains multiple repetitive elements."
     gene            4322. .31759
                     /gene="XRCC2"
                     /note="X-ray repair cross-complementing DNA repair gene"
     CDS             join(4322. .4360,19619. .19700,31038. .31759)
                     /gene="XRCC2"
                     /note="X-ray repair cross-complementing DNA repair
                     protein"
                     /codon_start=1
                     /product="Xrcc2"
                     /protein_id="AAC05802.1"
                     /db_xref="GI:2961445"
                     /translation="MCSAFHRAESGTELLARLEGRSSLKEIEPNLFADEDSPVHGDIL
                     EFHGPEGTGKTEMLYHLTARCILPKSEGGLEVEVLFIDTDYHFDMLRLVTILEHRLSQ
                     SSEEIIKYCLGRFFLVYCSSSTHLLLTLYSLESMFCSHPSLCLLILDSLSAFYWIDRV
                     NGGESVNLQESTLRKCSQCLEKLVNDYRLVLFATTQTIMQKASSSSEEPSHASRRLCD
                     VDIDYRPYLCKAWQQLVKHRMFFSKQDDSQSSNQFSLVSRCLKSNSLKKHFFIIGESG
                     VEFC"
     repeat_region   5618. .5774
                     /rpt_family="HSATI"
```

```
     repeat_region   complement(5783. .6067)
                     /rpt_family="Alu"
     repeat_region   complement(6510. .6810)
                     /rpt_family="Alu"
     repeat_region   6948. .7208
                     /rpt_family="Alu"
     repeat_region   complement(8375. .8630)
                     /rpt_family="Alu"
     repeat_region   8887. .9129
                     /rpt_family="MER42"
     repeat_region   complement(9181. .9456)
                     /rpt_family="Alu"
     repeat_region   9515. .9795
                     /rpt_family="Alu"
     repeat_region   complement(10534. .10574)
                     /rpt_family="Alu"
     repeat_region   10709. .10968
                     /rpt_family="Alu"
     repeat_region   12008. .12076
                     /rpt_family="MER44C"
     repeat_region   12109. .12326
                     /rpt_family="MER44C"
     repeat_region   complement(12359. .12632)
                     /rpt_family="Alu"
     repeat_region   complement(12978. .13268)
                     /rpt_family="Alu"
     repeat_region   13495. .13780
                     /rpt_family="Alu"
     repeat_region   13900. .14183
                     /rpt_family="Alu"
     repeat_region   14641. .14913
                     /rpt_family="Alu"
     repeat_region   complement(15901. .16104)
                     /rpt_family="MLT2B2"
     repeat_region   complement(16576. .16844)
                     /rpt_family="Alu"
     repeat_region   17529. .17829
                     /rpt_family="Alu"
     repeat_region   18140. .18430
                     /rpt_family="Alu"
     repeat_region   complement(19259. .19503)
                     /rpt_family="Alu"
     misc_feature    complement(20710. .20887)
                     /note="BLASTX similarity to  (1088. .1146); match: 0.45,
                     score: 9.0e-06; database searched: nr; hypothetical
                     protein (L1H 3' region) - human"
     repeat_region   20942. .21055
                     /rpt_family="Alu"
     repeat_region   21253. .21526
                     /rpt_family="L1"
     repeat_region   21931. .22256
                     /rpt_family="Alu"
     repeat_region   22264. .22440
                     /rpt_family="Alu"
     repeat_region   22820. .23070
                     /rpt_family="Alu"
     repeat_region   complement(23153. .23436)
                     /rpt_family="Alu"
     repeat_region   complement(24031. .24148)
                     /rpt_family="Alu"
     repeat_region   complement(24205. .24387)
                     /rpt_family="L1"
     repeat_region   complement(24404. .24958)
                     /rpt_family="Alu"
     repeat_region   complement(26186. .26513)
                     /rpt_family="Alu"
     repeat_region   26621. .26930
                     /rpt_family="Alu"
     repeat_region   complement(27018. .27369)
                     /rpt_family="THE1"
     repeat_region   complement(28211. .28504)
                     /rpt_family="Alu"
     repeat_region   28597. .28774
                     /rpt_family="Alu"
     repeat_region   complement(28851. .29076)
                     /rpt_family="Alu"
     repeat_region   29170. .29452
                     /rpt_family="Alu"
     repeat_region   complement(29487. .29781)
                     /rpt_family="Alu"
     misc_feature    31037. .31372
                     /gene="XRCC2"
                     /note="BLASTN similarity to Y08837 (1. .336); match: 0.99,
                     score: 7.0e-130; database searched: nt; H.sapiens mRNA for
                     RAD51-like protein~Other overlapping matches:
                     BLASTX similarity to PID (93. .149, 206. .243); match: 0.43,
                     score: 9.7e-06; database searched: nr; (Y13144) Rad51
                     homologue [Trypanosoma brucei]~BLASTX similarity to
                     2262211 (101. .162); match: 0.43, score: 1.8e-06; database
                     searched: nr; (U92068) RecA-like protein [Mus musculus]"
     repeat_region   complement(31829. .32256)
                     /rpt_family="Alu"
     repeat_region   32533. .33613
                     /rpt_family="L1"
     misc_feature    32880. .33124
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: good, score: 67.000"
     repeat_region   33635. .33888
                     /rpt_family="Alu"
     repeat_region   33903. .34050
                     /rpt_family="L1"
     repeat_region   complement(34480. .34609)
                     /rpt_family="Alu"
     repeat_region   35004. .35121
                     /rpt_family="Alu"
     misc_feature    36278. .36522
                     /note="DDS similarity to AA255626 zs31f06.s1 NCI_CGAP_GCB1
                     Homo sapiens cDNA clone 686819 3' similar to contains Alu
                     repetitive element;contains element LTR6 repetitive
                     element ;(186. .427); 96% identity."
     repeat_region   complement(36396. .36646)
                     /rpt_family="Alu"
     repeat_region   36742. .37029
                     /rpt_family="Alu"

  Query Match             80.0%;  Score 16;  DB 9;  Length 76850;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      27146 AGGAGGAACAAGTCCC 27161


RESULT 29
AX646415/c
LOCUS       AX646415               84510 bp    DNA     linear   PAT 04-MAR-2003
DEFINITION  Sequence 607 from Patent EP1270724.
ACCESSION   AX646415
VERSION     AX646415.1  GI:28798796
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Asai,K., Akiyama,Y. and Aburatani,H.
  TITLE     Guanosine triphosphate-binding protein coupled receptors
  JOURNAL   Patent: EP 1270724-A 607 02-JAN-2003;
            National Institute of Advanced Industrial Science and Technology
            (JP) ; Center for Advanced Science and Technology Incubation, Ltd.
            (JP)
FEATURES             Location/Qualifiers
     source          1. .84510
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                 /db_xref="taxon:9606"
     CDS         join(201. .272,29879. .29999,45509. .45616,50871. .50974,
                 58541. .58638,60172. .60304,62427. .62580,71080. .71140,
                 72348. .72417,75647. .75738,77542. .77668,82828. .82869,
                 84170. .84310)
                 /note="unnamed protein product"
                 /codon_start=1
                 /protein_id="CAD69336.1"
                 /db_xref="GI:28798797"
                 /translation="MRPHLSPPLQQLLLPVLLACAAHSTGALPRLCDVLQVLWEEQDQ
                 CLQELSREQTGDLGTEQPVPGCEGMWDNISCWPSSVPGRMVEVECPRFLRMLTSRNGS
                 LFRNCTQDGWSETFPRPNLACGVNVNDSSNEKRHSYLLKLKVMYTVGYSSSLVMLLVA
                 LGILCAFRRLHCTRNYIHMHLFVSFILRALSNFIKDAVLFSSDDVTYCDAHRAGCKLV
                 MVLFQYCIMANYSWLLVEGLYLHTLLAISFFSERKYLQGFVAFGWGSPAIFVALWAIA
                 RHFLEDVGCWDINANASIWWIIRGPVILSILINFILFINILRILMRKLRTQETRGNEV
                 SHYKRLARSTLLLIPLFGIHYIVFAFSPEDAMEIQLFFELALGSFQGLVVAVLYCFLN
                 GEVQLEVQKKWQQWHLREFPLHPVASFSNSTKASHLEQSQGTCRTSII"
BASE COUNT    21253 a  19989 c  19504 g  23764 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 6;  Length 84510;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db      66595 CTTAGGAGGAACAAGT 66580


RESULT 30
AB065660/c
LOCUS       AB065660               84510 bp    DNA     linear   PRI 23-JUL-2002
DEFINITION  Homo sapiens gene for seven transmembrane helix receptor, complete
            cds, isolate:CBRC7TM_223.
ACCESSION   AB065660
VERSION     AB065660.1  GI:21928594
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Sato,T., Okouchi,I., Arita,M., Futami,K., Matsumoto,S.,
            Tsutsumi,S., Aburatani,H., Asai,K. and Akiyama,Y.
  TITLE     Genome-wide discovery and analysis of human seven transmembrane
            helix receptor genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 84510)
  AUTHORS   Suwa,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research
            Center (CBRC), National Institute of Advanced Industrial Science
            and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan
            (E-mail:m-suwa@aist.go.jp, URL:http://www.cbrc.jp/,
            Tel:81-3-3599-8080, Fax:81-3-3599-8081)
COMMENT     This sequence is a seven transmembrane helix receptor candidate
            predicted from the whole human genome sequences using our automated
            system that contains programs of gene
            finding(GeneDecoder), sequence search, motif-domain assignment and
            transmembrane helix prediction.
            And the sequence is submitted by the collaborative project between
            [Computational Biology Research Center (CBRC), National Institute
            of Advanced Industrial Science and Technology (AIST)] and [Genome
            Science Division, Research Center for Advanced Science and
            Technology (RCAST), University of Tokyo].
FEATURES             Location/Qualifiers
     source          1. .84510
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="CBRC7TM_223"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     CDS             join(201. .272,29879. .29999,45509. .45616,50871. .50974,
                     58541. .58638,60172. .60304,62427. .62580,71080. .71140,
                     72348. .72417,75647. .75738,77542. .77668,82828. .82869,
                     84170. .84310)
                     /codon_start=1
                     /evidence=not_experimental
                     /product="seven transmembrane helix receptor"
                     /protein_id="BAC05886.1"
                     /db_xref="GI:21928595"
                     /translation="MRPHLSPPLQQLLLPVLLACAAHSTGALPRLCDVLQVLWEEQDQ
                     CLQELSREQTGDLGTEQPVPGCEGMWDNISCWPSSVPGRMVEVECPRFLRMLTSRNGS
                     LFRNCTQDGWSETFPRPNLACGVNVNDSSNEKRHSYLLKLKVMYTVGYSSSLVMLLVA
                     LGILCAFRRLHCTRNYIHMHLFVSFILRALSNFIKDAVLFSSDDVTYCDAHRAGCKLV
                     MVLFQYCIMANYSWLLVEGLYLHTLLAISFFSERKYLQGFVAFGWGSPAIFVALWAIA
                     RHFLEDVGCWDINANASIWWIIRGPVILSILINFILFINILRILMRKLRTQETRGNEV
                     SHYKRLARSTLLLIPLFGIHYIVFAFSPEDAMEIQLFFELALGSFQGLVVAVLYCFLN
                     GEVQLEVQKKWQQWHLREFPLHPVASFSNSTKASHLEQSQGTCRTSII"
BASE COUNT    21253 a  19989 c  19504 g  23764 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 84510;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db      66595 CTTAGGAGGAACAAGT 66580


RESULT 31
AC010963
LOCUS       AC010963               94357 bp    DNA     linear   HTG 01-OCT-1999
DEFINITION  Homo sapiens clone 11_J_20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC010963
VERSION     AC010963.2  GI:6006126
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone 11_J_20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94357)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 1, 1999 this sequence version replaced gi:5931399.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
            * NOTE: This record contains 98 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
```

```
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      892: contig of 892 bp in length
*                    gap of unknown length
*      893     1775: contig of 883 bp in length
*                    gap of unknown length
*     1776     2661: contig of 886 bp in length
*                    gap of unknown length
*     2662     3541: contig of 880 bp in length
*                    gap of unknown length
*     3542     4409: contig of 868 bp in length
*                    gap of unknown length
*     4410     5262: contig of 853 bp in length
*                    gap of unknown length
*     5263     6130: contig of 868 bp in length
*                    gap of unknown length
*     6131     7001: contig of 871 bp in length
*                    gap of unknown length
*     7002     7886: contig of 885 bp in length
*                    gap of unknown length
*     7887     8736: contig of 850 bp in length
*                    gap of unknown length
*     8737     9619: contig of 883 bp in length
*                    gap of unknown length
*     9620    10511: contig of 892 bp in length
*                    gap of unknown length
*    10512    11388: contig of 877 bp in length
*                    gap of unknown length
*    11389    12279: contig of 891 bp in length
*                    gap of unknown length
*    12280    13165: contig of 886 bp in length
*                    gap of unknown length
*    13166    14075: contig of 910 bp in length
*                    gap of unknown length
*    14076    14942: contig of 867 bp in length
*                    gap of unknown length
*    14943    15818: contig of 876 bp in length
*                    gap of unknown length
*    15819    16711: contig of 893 bp in length
*                    gap of unknown length
*    16712    17653: contig of 942 bp in length
*                    gap of unknown length
*    17654    18496: contig of 843 bp in length
*                    gap of unknown length
*    18497    19400: contig of 904 bp in length
*                    gap of unknown length
*    19401    20249: contig of 849 bp in length
*                    gap of unknown length
*    20250    21108: contig of 859 bp in length
*                    gap of unknown length
*    21109    21965: contig of 857 bp in length
*                    gap of unknown length
*    21966    22898: contig of 933 bp in length
*                    gap of unknown length
*    22899    23813: contig of 915 bp in length
*                    gap of unknown length
*    23814    24754: contig of 941 bp in length
*                    gap of unknown length
*    24755    25644: contig of 890 bp in length
*                    gap of unknown length
*    25645    26516: contig of 872 bp in length
*                    gap of unknown length
*    26517    27470: contig of 954 bp in length
*                    gap of unknown length
*    27471    28376: contig of 906 bp in length
*                    gap of unknown length
*    28377    29249: contig of 873 bp in length
*                    gap of unknown length
*    29250    30166: contig of 917 bp in length
*                    gap of unknown length
*    30167    31093: contig of 927 bp in length
*                    gap of unknown length
*    31094    32006: contig of 913 bp in length
*                    gap of unknown length
*    32007    32914: contig of 908 bp in length
*                    gap of unknown length
*    32915    33811: contig of 897 bp in length
*                    gap of unknown length
*    33812    34686: contig of 875 bp in length
*                    gap of unknown length
*    34687    35582: contig of 896 bp in length
*                    gap of unknown length
*    35583    36464: contig of 882 bp in length
*                    gap of unknown length
*    36465    37359: contig of 895 bp in length
*                    gap of unknown length
*    37360    38314: contig of 955 bp in length
*                    gap of unknown length
*    38315    39271: contig of 957 bp in length
*                    gap of unknown length
*    39272    40263: contig of 992 bp in length
*                    gap of unknown length
*    40264    41307: contig of 1044 bp in length
*                    gap of unknown length
*    41308    42294: contig of 987 bp in length
*                    gap of unknown length
*    42295    43308: contig of 1014 bp in length
*                    gap of unknown length
*    43309    44254: contig of 946 bp in length
*                    gap of unknown length
*    44255    45242: contig of 988 bp in length
*                    gap of unknown length
*    45243    46294: contig of 1052 bp in length
*                    gap of unknown length
*    46295    47278: contig of 984 bp in length
*                    gap of unknown length
*    47279    48315: contig of 1037 bp in length
*                    gap of unknown length
*    48316    49271: contig of 956 bp in length
*                    gap of unknown length
*    49272    50286: contig of 1015 bp in length
*                    gap of unknown length
*    50287    51235: contig of 949 bp in length
*                    gap of unknown length
*    51236    52294: contig of 1059 bp in length
*                    gap of unknown length
*    52295    53307: contig of 1013 bp in length
*                    gap of unknown length
*    53308    54384: contig of 1077 bp in length
*                    gap of unknown length
*    54385    55433: contig of 1049 bp in length
*                    gap of unknown length
*    55434    56429: contig of 996 bp in length
*                    gap of unknown length
*    56430    57414: contig of 985 bp in length
*                    gap of unknown length
*    57415    58419: contig of 1005 bp in length
*                    gap of unknown length
*    58420    59378: contig of 959 bp in length
*                    gap of unknown length
*    59379    60349: contig of 971 bp in length
*                    gap of unknown length
*    60350    61324: contig of 975 bp in length
*                    gap of unknown length
*    61325    62374: contig of 1050 bp in length
*                    gap of unknown length
*    62375    63342: contig of 968 bp in length
*                    gap of unknown length
*    63343    64279: contig of 937 bp in length
*                    gap of unknown length
*    64280    65353: contig of 1074 bp in length
*                    gap of unknown length
*    65354    66383: contig of 1030 bp in length
*                    gap of unknown length
*    66384    67403: contig of 1020 bp in length
```

```
            *                    gap of unknown length
            *    67404    68422: contig of 1019 bp in length
            *                    gap of unknown length
            *    68423    69481: contig of 1059 bp in length
            *                    gap of unknown length
            *    69482    70481: contig of 1000 bp in length
            *                    gap of unknown length
            *    70482    71504: contig of 1023 bp in length
            *                    gap of unknown length
            *    71505    72501: contig of 997 bp in length
            *                    gap of unknown length

  Query Match             80.0%;  Score 16;  DB 2;  Length 94357;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      61206 AGGAGGAACAAGTCCC 61221


RESULT 32
AP000032/c
LOCUS       AP000032              100000 bp    DNA     linear   PRI 20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, segment 3/28,
            complete sequence.
ACCESSION   AP000032
VERSION     AP000032.1  GI:3132342
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA, chromosome 21q
  JOURNAL   Published Only in DataBase (1998)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Masahira Hattori, Kitasato University,
            Department of Science, JST Sequencing Laboratory; Kitasato 1-15-1,
            Sagamihara 228, Japan (E-mail:hattori@hgc.ims.u-tokyo.ac.jp,
            Tel:0427-78-9732, Fax:0427-78-9561)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by:Human Genome Sequencing in ALIS project of
            JST
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top. html)
            or send email to webmaster@www-alis.tokyo.jst.go.jp'.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="S306-f74A4f.seq_1"
BASE COUNT    27721 a  22808 c  23582 g  25889 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 100000;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      11338 AGGAGGAACAAGTCCC 11323


RESULT 33
AP000104
LOCUS       AP000104              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
            Q78C10-149C3 region, segment 7/20.
ACCESSION   AP000104
VERSION     AP000104.1  GI:4730838
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 2,051,516bp genomic DNA of 21q22.1 GART and AML region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
     STS             3265. .3414
                     /note="SHGC-37672;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:6458696"
     STS             3576. .3747
                     /standard_name="D21S1413"
                     /note="UT7582;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:315533"
     STS             11676. .11835
                     /note="SHGC-51923;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:6464667"
     STS             11723. .11855
                     /standard_name="D21S224"
                     /note="KM838/KM839;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:182624"
     STS             44121. .44250
                     /note="SGC35207;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:4581832"
     STS             44275. .44383
```

```
                /note="SHGC-51895;The location is between each flanking
                site of PCR primers."
                /db_xref="GDB:6464619"
     STS        46239. .46473
                /standard_name="D21S1879"
                /note="PrG-13419/PrG-13420;The location is between each
                flanking site of PCR primers."
                /db_xref="GDB:598243"
     STS        53587. .53745
                /standard_name="D21S1693"
                /note="134H10aR-F/134H10aR-R;The location is between each
                flanking site of PCR primers."
                /db_xref="GDB:438493"
     STS        82233. .82396
                /note="SHGC-51815;The location is between each flanking
                site of PCR primers."
                /db_xref="GDB:6464271"
BASE COUNT    27206 a  22604 c  22308 g  27882 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 100000;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      25561 AGGAGGAACAAGTCCC 25576


RESULT 34
AP000180
LOCUS       AP000180              100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone Q78C10-f32E9, segment 7/21, complete sequence.
ACCESSION   AP000180
VERSION     AP000180.1  GI:4827079
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 2,083,744bp genomic DNA of 21q22.1 (REGION:
            D21S226-AML CLONE RANGE: Q78C10-f32E9)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
            original/previous sequences.  We determined the boundary between
            the insertion and genomic sequences experimentally, removed the
            insertion sequences, reconstituted the present data. The sequencing
            project is supported by Japan Science Technology Corporation (JST)
            and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
BASE COUNT    27206 a  22604 c  22308 g  27882 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 100000;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db      25561 AGGAGGAACAAGTCCC 25576


RESULT 35
AP000681
LOCUS       AP000681              109149 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-2L13 map 11q23, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AP000681
VERSION     AP000681.3  GI:8118869
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109149)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 109,149 genomic DNA of 11q23
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 109149)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997555.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-2L13
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 99833 bases at least Q40
                Consensus quality: 104437 bases at least Q30
                Consensus quality: 106696 bases at least Q20
                Insert size: 107949; sum-of-contigs
                Quality coverage: 4.90x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            13 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                    1     15418 contig of  15418 bp in length
                15519     32589 contig of  17071 bp in length
                32690     46533 contig of  13844 bp in length
                46634     59641 contig of  13008 bp in length
                59742     69969 contig of  10228 bp in length
                70070     79400 contig of   9331 bp in length
                79501     86689 contig of   7189 bp in length
                86790     94162 contig of   7373 bp in length
                94263    100141 contig of   5879 bp in length
               100242    104375 contig of   4134 bp in length
```

```
            104476    107052 contig of   2577 bp in length
            107153    108671 contig of   1519 bp in length
            108772    109149 contig of    378 bp in length
          Sequence updated (01-Feb-2000)
          Sequence updated (26-May-2000).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    15418: contig of 15418 bp in length
          *    15419    15518: gap of 100 bp
          *    15519    32589: contig of 17071 bp in length
          *    32590    32689: gap of 100 bp
          *    32690    46533: contig of 13844 bp in length
          *    46534    46633: gap of 100 bp
          *    46634    59641: contig of 13008 bp in length
          *    59642    59741: gap of 100 bp
          *    59742    69969: contig of 10228 bp in length
          *    69970    70069: gap of 100 bp
          *    70070    79400: contig of 9331 bp in length
          *    79401    79500: gap of 100 bp
          *    79501    86689: contig of 7189 bp in length
          *    86690    86789: gap of 100 bp
          *    86790    94162: contig of 7373 bp in length
          *    94163    94262: gap of 100 bp
          *    94263   100141: contig of 5879 bp in length
          *   100142   100241: gap of 100 bp
          *   100242   104375: contig of 4134 bp in length
          *   104376   104475: gap of 100 bp
          *   104476   107052: contig of 2577 bp in length
          *   107053   107152: gap of 100 bp
          *   107153   108671: contig of 1519 bp in length
          *   108672   108771: gap of 100 bp
          *   108772   109149: contig of 378 bp in length.
FEATURES             Location/Qualifiers
     source          1. .109149
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q23"
                     /clone="CMB9-2L13"
     misc_feature    1. .15418
                     /note="assembly_fragment"
     misc_feature    15519. .32589
                     /note="assembly_fragment"
     misc_feature    32690. .46533
                     /note="assembly_fragment"
     misc_feature    46634. .59641
                     /note="assembly_fragment"
     misc_feature    59742. .69969
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    70070. .79400
                     /note="assembly_fragment"
     misc_feature    79501. .86689
                     /note="assembly_fragment"
     misc_feature    86790. .94162
                     /note="assembly_fragment"
     misc_feature    94263. .100141
                     /note="assembly_fragment"
     misc_feature    100242. .104375
                     /note="assembly_fragment"
     misc_feature    104476. .107052
                     /note="assembly_fragment"
     misc_feature    107153. .108671
                     /note="assembly_fragment"
     misc_feature    108772. .109149
                     /note="assembly_fragment clone_end:T7 vector_side:right"
BASE COUNT    32146 a  20393 c  21206 g  34199 t   1205 others
```

```
ORIGIN

  Query Match           80.0%;  Score 16;  DB 2;  Length 109149;
  Best Local Similarity  100.0%;  Pred. No. 79;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db      73711 CTTAGGAGGAACAAGT 73726


Search completed: August 15, 2003, 09:33:56
Job time : 555.75 secs
```

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:41:37 ; Search time 1252 Seconds
(without alignments)
388.250 Million cell updates/sec

Title: US-10-074-620-2
Perfect score: 20
Sequence: 1 ccttaggaggaacaagtccc 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		17	85.0	676	10	BE964731	BE964731 601658274
2	c	17	85.0	817	10	BF142291	BF142291 601791832
3		16	80.0	200	12	BM751719	BM751719 K-EST0027
4		16	80.0	203	10	BE755758	BE755758 209682 MA
5		16	80.0	242	14	Z42603	Z42603 HSC0PA091 n
6		16	80.0	350	9	AU231688	AU231688 AU231688
7		16	80.0	358	14	R24852	R24852 yg29e05.r1
8		16	80.0	390	28	AQ483201	AQ483201 RPCI-11-2
9		16	80.0	395	9	AI696726	AI696726 wc56d11.x
10		16	80.0	400	14	T75305	T75305 yc89c08.r1
11		16	80.0	413	28	AQ379270	AQ379270 RPCI11-15
12		16	80.0	431	9	AW769407	AW769407 hl63g12.x
13		16	80.0	443	10	BE227333	BE227333 894030E10
14		16	80.0	464	13	BY414889	BY414889 BY414889
15		16	80.0	476	14	R24326	R24326 yg32f03.r1
16	c	16	80.0	487	10	BF830043	BF830043 MR3-HN005
17		16	80.0	494	14	R24340	R24340 yg32h03.r1
18	c	16	80.0	516	10	BB693705	BB693705 BB693705
19	c	16	80.0	516	12	BG994855	BG994855 PM0-HT116
20		16	80.0	524	14	CA405555	CA405555 1001704 H
21		16	80.0	550	13	BQ347567	BQ347567 CM0-HT017
22		16	80.0	554	14	W52429	W52429 zc94e08.r1
23	c	16	80.0	591	12	BI751017	BI751017 Ta01_05d0
24		16	80.0	599	2	HSM095392	Bx501675 Homo sapi
25		16	80.0	615	9	AU137185	AU137185 AU137185
26		16	80.0	630	12	BI156614	BI156614 602921206
27		16	80.0	661	9	AV821593	AV821593 AV821593
28	c	16	80.0	668	12	BJ273261	BJ273261 BJ273261
29		16	80.0	703	13	BQ539794	BQ539794 PTAM0102
30		16	80.0	740	13	BU635192	BU635192 003D06 In
31		16	80.0	745	12	BI115344	BI115344 602863159
32		16	80.0	871	10	BE869673	BE869673 601445775
33		16	80.0	872	12	BI088175	BI088175 602851213
34		16	80.0	879	13	BU177008	BU177008 AGENCOURT
35		16	80.0	884	10	BG740602	BG740602 602631028
36		16	80.0	886	10	BG501889	BG501889 602548991
37		16	80.0	900	13	BU509687	BU509687 AGENCOURT
38		16	80.0	904	10	BE869754	BE869754 601445669
39		16	80.0	942	13	BQ225365	BQ225365 AGENCOURT
40		16	80.0	966	13	BQ277316	BQ277316 AGENCOURT
41		16	80.0	978	9	AL536291	AL536291 AL536291
42		16	80.0	1201	9	AL546318	AL546318 AL546318
43	c	15	75.0	179	13	BY607541	BY607541 BY607541
44		15	75.0	193	28	BZ181647	BZ181647 CH230-340
45	c	15	75.0	205	28	B59389	B59389 CIT-HSP-201
46		15	75.0	280	10	BE055057	BE055057 GA__Ea003
47		15	75.0	286	10	BB450936	BB450936 BB450936
48		15	75.0	287	10	BE054889	BE054889 GA__Ea002
49		15	75.0	287	13	BQ403153	BQ403153 GA__Ed005
50	c	15	75.0	287	13	BQ412134	BQ412134 GA__Ed005
51		15	75.0	307	28	BH071600	BH071600 RPCI-24-2
52		15	75.0	315	10	BB317883	BB317883 BB317883
53	c	15	75.0	372	9	AA155221	AA155221 mr97c03.r
54	c	15	75.0	372	9	AA155224	AA155224 mr97d03.r
55	c	15	75.0	380	9	AA155220	AA155220 mr97c02.r
56		15	75.0	387	13	BY155062	BY155062 BY155062
57	c	15	75.0	413	9	AA155222	AA155222 mr97c04.r
58		15	75.0	421	28	AZ318410	AZ318410 1M0037P14
59	c	15	75.0	437	9	AA546043	AA546043 vk61c08.r
60		15	75.0	439	28	BH705996	BH705996 BOMDK80TR
61		15	75.0	452	28	AQ759784	AQ759784 HS_2257_A
62	c	15	75.0	453	9	AI333060	AI333060 qq17b11.x
63		15	75.0	460	14	N80106	N80106 yz87f02.r1
64		15	75.0	460	14	N80119	N80119 yz87h02.r1
65		15	75.0	460	28	AZ802618	AZ802618 2M0061O03
66	c	15	75.0	475	28	AZ943693	AZ943693 2M0204O22
67		15	75.0	477	10	BF270306	BF270306 GA__Eb000
68		15	75.0	481	10	BF957102	BF957102 QV4-NN114
69		15	75.0	487	13	BQ411430	BQ411430 GA__Ed003
70		15	75.0	489	28	BH076127	BH076127 RPCI-24-2
71	c	15	75.0	494	10	BE721743	BE721743 189459 MA
72	c	15	75.0	496	28	AZ766878	AZ766878 1M0564A15
73	c	15	75.0	499	29	BZ925689	BZ925689 CH240_72D
74	c	15	75.0	505	28	AZ597111	AZ597111 1M0410I13
75	c	15	75.0	510	9	AJ225437	AJ225437 AJ225437
76		15	75.0	519	9	AV856135	AV856135 AV856135
77		15	75.0	523	28	AZ015440	AZ015440 RPCI-23-3

	78	15	75.0	524	10	BG440460 BG440460 GA__Ea000
c	79	15	75.0	537	28	AQ951596 AQ951596 Sheared D
c	80	15	75.0	538	28	AQ558274 AQ558274 HS_2086_B
c	81	15	75.0	542	14	CA873512 CA873512 K0925G06-
	82	15	75.0	555	13	BQ405699 BQ405699 GA__Ed008
c	83	15	75.0	555	13	BQ414368 BQ414368 GA__Ed008
	84	15	75.0	556	28	AZ079907 AZ079907 RPCI-23-4
	85	15	75.0	560	28	AZ625916 1M0465C24
	86	15	75.0	562	28	AQ788800 AQ788800 HS_3187_B
c	87	15	75.0	566	28	AZ792636 AZ792636 2M0045C15
	88	15	75.0	573	28	AZ625800 1M0465I12
	89	15	75.0	575	9	AV880832 AV880832 AV880832
	90	15	75.0	577	13	BU694428 BU694428 LL2in1197
	91	15	75.0	582	29	BZ886748 BZ886748 CH240_269
	92	15	75.0	592	28	AQ462877 AQ462877 HS_5212_A
c	93	15	75.0	597	10	BE402486 BE402486 CSB008D12
c	94	15	75.0	597	13	BQ608042 BQ608042 BRY_3943
	95	15	75.0	601	9	AA422164 AA422164 zv31c01.r
	96	15	75.0	606	10	BE288544 BE288544 601094160
c	97	15	75.0	607	29	AG150407 AG150407 Pan trogl
c	98	15	75.0	609	9	AI927904 AI927904 wp03c07.x
	99	15	75.0	611	29	BX133940 BX133940 Danio rer
	100	15	75.0	617	10	BG441695 BG441695 GA__Ea001
	101	15	75.0	617	28	AZ261932 AZ261932 RPCI-23-1
	102	15	75.0	620	10	BG442675 BG442675 GA__Ea001
	103	15	75.0	626	9	AV880047 AV880047 AV880047
	104	15	75.0	626	28	BH059230 BH059230 RPCI-24-3
	105	15	75.0	628	13	BQ410892 BQ410892 GA__Ed003
c	106	15	75.0	630	13	BQ413572 BQ413572 GA__Ed007
c	107	15	75.0	634	10	BE985077 BE985077 UI-M-CG0p
c	108	15	75.0	634	13	BQ410893 BQ410893 GA__Ed003
	109	15	75.0	636	10	BG442578 BG442578 GA__Ea001
c	110	15	75.0	640	13	BQ407961 BQ407961 GA__Ed000
c	111	15	75.0	640	13	BQ414737 BQ414737 GA__Ed009
	112	15	75.0	641	28	AZ113409 AZ113409 RPCI-23-4
c	113	15	75.0	643	9	AI585535 AI585535 vk61c08.y
	114	15	75.0	645	9	AV889725 AV889725 AV889725
c	115	15	75.0	646	9	AV856437 AV856437 AV856437
c	116	15	75.0	646	28	AZ255189 AZ255189 RPCI-23-1
	117	15	75.0	650	28	AZ458615 AZ458615 1M0262E23
c	118	15	75.0	651	13	BQ413525 BQ413525 GA__Ed007
	119	15	75.0	652	10	BG442693 BG442693 GA__Ea001
c	120	15	75.0	653	10	BG072204 BG072204 H3107H11-

ALIGNMENTS

```
RESULT 1
BE964731
LOCUS       BE964731                 676 bp    mRNA    linear   EST 14-DEC-2000
DEFINITION  601658274R1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885642 3',
            mRNA sequence.
ACCESSION   BE964731
VERSION     BE964731.2  GI:11768351
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     On Oct 3, 2000 this sequence version replaced gi:10575436.
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM648  row: c  column: 19
            High quality sequence stop: 227.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3885642"
                     /tissue_type="large cell carcinoma, undifferentiated"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_69"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      139 a    138 c    193 g    206 t
ORIGIN

  Query Match             85.0%;  Score 17;  DB 10;  Length 676;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAGTCC 19
              |||||||||||||||||
Db        436 TTAGGAGGAACAAGTCC 452


RESULT 2
BF142291/c
LOCUS       BF142291                 817 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  601791832F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4022647 5',
            mRNA sequence.
ACCESSION   BF142291
VERSION     BF142291.1  GI:10981241
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9279  row: p  column: 08
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4022647"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu30"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally.  Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      186 a    237 c    201 g    193 t
ORIGIN
```

```
  Query Match            85.0%;  Score 17;  DB 10;  Length 817;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCTTAGGAGGAACAAGT 17
              |||||||||||||||||
Db        773 CCTTAGGAGGAACAAGT 757


RESULT 3
BM751719
LOCUS       BM751719                 200 bp    mRNA    linear   EST 04-MAR-2002
DEFINITION  K-EST0027999 S9SNU601 Homo sapiens cDNA clone S9SNU601-24-D05 5',
            mRNA sequence.
ACCESSION   BM751719
VERSION     BM751719.1  GI:19081351
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 24  row: D  column: 05
            High quality sequence stop: 200.
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S9SNU601-24-D05"
                     /sex="M"
                     /tissue_type="Ascites"
                     /cell_type="Epithelial"
                     /cell_line="SNU-601"
                     /lab_host="Top10F'"
                     /clone_lib="S9SNU601"
                     /note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
                     Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including SfiI
                     site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized with Superscript II using SfiI
                     oligo-dT primer. After first strand synthesis, RNA was
                     degraded by NaOH treatment and cDNA was amplified by PCR
                     reaction. The PCR products were digested with SfiI and
                     cloned into DraIII- digested pME18S-FL3 vector. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT       73 a     34 c     37 g     56 t
ORIGIN

  Query Match            80.0%;  Score 16;  DB 12;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         92 CTTAGGAGGAACAAGT 107


RESULT 4
BE755758
LOCUS       BE755758                 203 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  209682 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE755758
VERSION     BE755758.1  GI:10169750
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 203)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
   PUBMED   11282978
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 59  row: L  column: 7
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .203
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       70 a     29 c     48 g     56 t
ORIGIN

  Query Match            80.0%;  Score 16;  DB 10;  Length 203;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        102 CTTAGGAGGAACAAGT 117


RESULT 5
Z42603
LOCUS       Z42603                   242 bp    mRNA    linear   EST 10-NOV-1994
DEFINITION  HSC0PA091 normalized infant brain cDNA Homo sapiens cDNA clone
            c-0pa09, mRNA sequence.
ACCESSION   Z42603
VERSION     Z42603.1  GI:567353
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
   PUBMED   7757816
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-0pa09
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="c-0pa09"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /clone_lib="normalized infant brain cDNA"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total brain
                     ;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       88 a     42 c     43 g     69 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 14;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         48 CTTAGGAGGAACAAGT 63


RESULT 6
AU231688
LOCUS       AU231688                350 bp    mRNA    linear   EST 21-SEP-2001
DEFINITION  AU231688 Cloned bovine fetus cDNA Bos taurus cDNA clone Cln595 3',
            mRNA sequence.
ACCESSION   AU231688
VERSION     AU231688.1  GI:15719980
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   Taniguchi,Y., Lejukole,H.Y., Yamada,T., Akagi,S., Yasue,H. and
            Sasaki,Y.
  TITLE     Analysis of expressed sequence tags from a cDNA library of somatic
            nuclear transfer-derived cloned bovine fetus
  JOURNAL   Unpublished
COMMENT     Contact: Takahisa Yamada
            Graduate School of Agriculture
            Kyoto University
            Sakyoku, Kitashirakawa, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-6323
            Fax: 81-75-753-6340
            Email: tyamada@jkans.jkans.kais.kyoto-u.ac.j
            This clone was obtained from a 3' end cDNA library.
FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="Cln595"
                     /dev_stage="fetus"
                     /clone_lib="Cloned bovine fetus cDNA"
BASE COUNT      101 a     70 c     83 g     92 t      4 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 350;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        252 CTTAGGAGGAACAAGT 267


RESULT 7
R24852
LOCUS       R24852                  358 bp    mRNA    linear   EST 20-APR-1995
DEFINITION  yg29e05.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:33804 5', mRNA sequence.
ACCESSION   R24852
VERSION     R24852.1  GI:779740
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2215
            High quality sequence stops: 284 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2215   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 284.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:406151"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:33804"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'];
```

double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 112 a 64 c 68 g 112 t 2 others
ORIGIN

Query Match 80.0%; Score 16; DB 14; Length 358;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 47 CTTAGGAGGAACAAGT 62

RESULT 8
AQ483201
LOCUS AQ483201 390 bp DNA linear GSS 24-APR-1999
DEFINITION RPCI-11-241P16.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-241P16, genomic survey sequence.
ACCESSION AQ483201
VERSION AQ483201.1 GI:4670605
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-11-241P16.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .390
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7592535"
/db_xref="taxon:9606"
/clone="RPCI-11-241P16"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 112 a 85 c 89 g 104 t
ORIGIN

Query Match 80.0%; Score 16; DB 28; Length 390;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 289 AGGAGGAACAAGTCCC 304

RESULT 9
AI696726
LOCUS AI696726 395 bp mRNA linear EST 17-DEC-1999
DEFINITION wc56d11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2322645 3' similar to contains OFR.t1 OFR OFR repetitive element ;, mRNA sequence.
ACCESSION AI696726
VERSION AI696726.1 GI:4984626
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 395)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 551 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 394.
FEATURES Location/Qualifiers
source 1. .395
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2322645"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pr28"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 125 a 85 c 97 g 88 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 84 AGGAGGAACAAGTCCC 99

RESULT 10
T75305
LOCUS T75305 400 bp mRNA linear EST 03-MAR-1995
DEFINITION yc89c08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:23341 5', mRNA sequence.
ACCESSION T75305
VERSION T75305.1 GI:692067
KEYWORDS EST.
SOURCE Homo sapiens (human)

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2260
            High quality sequence stops: 318 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2260   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 318.
FEATURES              Location/Qualifiers
     source           1. .400
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:395688"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:23341"
                      /sex="female"
                      /dev_stage="73 days post natal"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares infant brain 1NIB"
                      /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                      I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                      I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']
                      double-stranded cDNA was ligated to Hind III adaptors
                      (Pharmacia), digested with Not I and directionally cloned
                      into the Not I and Hind III sites of the Lafmid BA vector.
                      Library went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      123 a     77 c     72 g    127 t      1 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 14;  Length 400;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         26 CTTAGGAGGAACAAGT 41


RESULT 11
AQ379270
LOCUS       AQ379270                413 bp    DNA     linear   GSS 20-MAY-1999
DEFINITION  RPCI11-151H4.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-151H4,
            genomic survey sequence.
ACCESSION   AQ379270
VERSION     AQ379270.1  GI:4350293
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI11-151H4.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .413
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="GDB:7557771"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-151H4"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /clone_lib="RPCI-11"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      124 a     92 c     92 g    104 t      1 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 28;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        273 AGGAGGAACAAGTCCC 288


RESULT 12
AW769407
LOCUS       AW769407                431 bp    mRNA    linear   EST 04-MAY-2000
DEFINITION  hl63g12.x1 NCI_CGAP_Kid13 Homo sapiens cDNA clone IMAGE:3005926 3'
            similar to contains OFR.b1 OFR repetitive element ;, mRNA sequence.
ACCESSION   AW769407
VERSION     AW769407.1  GI:7701438
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 396.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Homo sapiens"
```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3005926"
/tissue_type="2 pooled Wilms' tumors, one primary and one metastatic to brain"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid13"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."
BASE COUNT 131 a 98 c 103 g 99 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 431;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 89 AGGAGGAACAAGTCCC 104

RESULT 13
BE227333
LOCUS BE227333 443 bp mRNA linear EST 06-JUL-2000
DEFINITION 894030E10.x3 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BE227333
VERSION BE227333.1 GI:8932572
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 443)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished
COMMENT Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.
FEATURES Location/Qualifiers
source 1. .443
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 89 a 129 c 139 g 86 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 443;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AGGAGGAACAAGTCCC 20
||||||||||||||||
Db 29 AGGAGGAACAAGTCCC 44

RESULT 14
BY414889
LOCUS BY414889 464 bp mRNA linear EST 13-DEC-2002
DEFINITION BY414889 RIKEN full-length enriched, 16 days embryo kidney Mus musculus cDNA clone I920006J23 3', mRNA sequence.
ACCESSION BY414889
VERSION BY414889.1 GI:26679833
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 464)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I920006J23"
                     /tissue_type="kidney"
                     /dev_stage="16 days embryo"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     kidney"
BASE COUNT      132 a     84 c    101 g    146 t      1 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 13;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        241 AGGAGGAACAAGTCCC 256


RESULT 15
R24326
LOCUS       R24326                  476 bp    mRNA    linear   EST 20-APR-1995
DEFINITION  yg32f03.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:33987 5', mRNA sequence.
ACCESSION   R24326
VERSION     R24326.1  GI:779214
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 349
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13RP1
            High quality sequence stop: 349.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:406334"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:33987"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      154 a     87 c     86 g    148 t      1 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 14;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         47 CTTAGGAGGAACAAGT 62


RESULT 16
BF830043/c
LOCUS       BF830043                487 bp    mRNA    linear   EST 13-JAN-2001
DEFINITION  MR3-HN0052-271200-001-b01 HN0052 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF830043
VERSION     BF830043.1  GI:12176257
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-HN0052-
            271200-001-b01&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 32
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HN0052"
                     /note="Organ: head_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
```

No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 165 a 102 c 108 g 112 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 487;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 483 CTTAGGAGGAACAAGT 468

RESULT 17
R24340
LOCUS R24340 494 bp mRNA linear EST 20-APR-1995
DEFINITION yg32h03.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
IMAGE:34200 5', mRNA sequence.
ACCESSION R24340
VERSION R24340.1 GI:779228
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2185
High quality sequence stops: 325 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2185 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 325.
FEATURES Location/Qualifiers
source 1. .494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:406547"
/db_xref="taxon:9606"
/clone="IMAGE:34200"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain 1NIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 151 a 90 c 90 g 159 t 4 others
ORIGIN

Query Match 80.0%; Score 16; DB 14; Length 494;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 47 CTTAGGAGGAACAAGT 62

RESULT 18
BB693705/c
LOCUS BB693705 516 bp mRNA linear EST 10-OCT-2001
DEFINITION BB693705 RIKEN full-length enriched, 2 days neonate sympathetic
ganglion Mus musculus cDNA clone 7120448G11 3', mRNA sequence.
ACCESSION BB693705
VERSION BB693705.1 GI:16020438
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 516)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
, Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
, Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
JOURNAL Unpublished
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .516
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7120448G11"
/sex="mixed"
/tissue_type="sympathetic ganglion"
/dev_stage="2 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 2 days neonate
sympathetic ganglion"

```
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I.
                     Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
                     ~RNA was provided by Akira Nakagawara, Div. of
                     Biochemistry, Chiba Cancer Center Research Institute,
                     666-2 Nitona, Chuoh-ku, Chiba, 260-8717 Japan, whose
                     assistance we gratefully acknowledge."
BASE COUNT      134 a    123 c    133 g    126 t
ORIGIN

  Query Match           80.0%;  Score 16;  DB 10;  Length 516;
  Best Local Similarity  100.0%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAGTC 18
              ||||||||||||||||
Db        375 TTAGGAGGAACAAGTC 360


RESULT 19
BG994855/c
LOCUS       BG994855                516 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  PM0-HT1166-150201-002-c08 HT1166 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG994855
VERSION     BG994855.1  GI:14398925
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-HT1166-
            150201-002-c08&t3=2001-02-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 516.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT1166"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      114 a    120 c    115 g    167 t
ORIGIN

  Query Match           80.0%;  Score 16;  DB 12;  Length 516;
  Best Local Similarity  100.0%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        414 AGGAGGAACAAGTCCC 399


RESULT 20
CA405555
LOCUS       CA405555                524 bp    mRNA    linear   EST 07-NOV-2002
DEFINITION  1001704 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CA405555
VERSION     CA405555.1  GI:24770426
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT      180 a     90 c     84 g    169 t      1 others
ORIGIN

  Query Match           80.0%;  Score 16;  DB 14;  Length 524;
  Best Local Similarity  100.0%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        161 CTTAGGAGGAACAAGT 176


RESULT 21
BQ347567
```

```
LOCUS       BQ347567                 550 bp    mRNA    linear   EST 20-MAY-2002
DEFINITION  CM0-HT0179-051099-064-g10 HT0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ347567
VERSION     BQ347567.1  GI:21011623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-HT0179-
            051099-064-g10&t3=1999-10-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 2.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0179"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      179 a    111 c     94 g    166 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 13;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        279 CTTAGGAGGAACAAGT 294


RESULT 22
W52429
LOCUS       W52429                   554 bp    mRNA    linear   EST 31-MAY-1996
DEFINITION  zc94e08.r1 Pancreatic Islet Homo sapiens cDNA clone IMAGE:338822
            5', mRNA sequence.
ACCESSION   W52429
VERSION     W52429.1  GI:1349780
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:1264196"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:338822"
                     /tissue_type="pancreatic islet"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /clone_lib="Pancreatic Islet"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
                     Takeda et al. Cloned unidirectionally. Primer: Oligo dT.
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      181 a     97 c     95 g    180 t      1 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 14;  Length 554;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        412 CTTAGGAGGAACAAGT 427


RESULT 23
BI751017/c
LOCUS       BI751017                 591 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  Ta01_05d02_C
            Ta01_AAFC_ECORC_Fusarium_graminearum_inoculated_wheat_heads
            Triticum aestivum cDNA clone Ta01_05d02, mRNA sequence.
ACCESSION   BI751017
VERSION     BI751017.1  GI:15772819
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Ouellet,T., Dan,H., Koul,A., Chapados,J., Couroux,P., De Moors,A.,
            Harris,L.J., Hattori,J.I., Robert,L.S., Singh,J.A, Sprott,D. and
            Tinker,N.A.
  TITLE     Expressed Sequence Tags from Wheat Heads 24 Hours after Spray
            Inoculation with Fusarium graminearum
  JOURNAL   Unpublished
COMMENT     Contact: Ouellet, Therese
            Eastern Cereal and Oilseed Research Centre
            Agriculture and Agri-food Canada
            Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
```

```
            CANADA
            Tel: (613) 759-1658
            Fax: (613) 759-1701
            Email: ouellettr@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Frontana"
                     /db_xref="taxon:4565"
                     /clone="Ta01_05d02"
                     /tissue_type="heads"
                     /dev_stage="anthesis"
                     /clone_lib="Ta01_AAFC_ECORC_Fusarium_graminearum_inoculate
                     d_wheat_heads"
                     /note="Vector: pGEM-T easy; Site_1: EcoRI; Site_2: EcoR1;
                     Controled chamber-grown wheat heads were spray inoculated
                     at mid-anthesis  with a Fusarium graminearum macroconidial
                     suspension (50,000 spores/ml)  and kept under intermitent
                     misting for 24 hours, then collected and  immediately
                     frozen in liquid nitrogen."
BASE COUNT      162 a    157 c    132 g    140 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 12;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAGTC 18
              ||||||||||||||||
Db        272 TTAGGAGGAACAAGTC 257


RESULT 24
HSM095392
ID   HSM095392  standard; RNA; EST; 599 BP.
XX
AC   BX501675;
XX
SV   BX501675.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp779F1664_r1  (from clone DKFZp779F1664)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-599
RA   Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA   Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
CC   consortium of the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp779F1664) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .599
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp779F1664"
FT                   /clone_lib="779 (synonym: hncc1). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="fetal"
FT                   /tissue_type="liver"
XX
SQ   Sequence 599 BP; 182 A; 105 C; 100 G; 212 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 2;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        479 CTTAGGAGGAACAAGT 494


RESULT 25
AU137185
LOCUS       AU137185                615 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU137185 PLACE1 Homo sapiens cDNA clone PLACE1005960 5', mRNA
            sequence.
ACCESSION   AU137185
VERSION     AU137185.1  GI:10997724
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1005960"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
BASE COUNT      200 a    124 c     92 g    196 t      3 others
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 615;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db         87 CTTAGGAGGAACAAGT 102


RESULT 26
BI156614
LOCUS       BI156614                630 bp    mRNA    linear   EST 05-JUL-2001
```

```
DEFINITION  602921206F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5061625 5',
            mRNA sequence.
ACCESSION   BI156614
VERSION     BI156614.1  GI:14616615
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11166  row: k  column: 02
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5061625"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT      145 a    133 c    166 g    186 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 12;  Length 630;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAGTC 18
              ||||||||||||||||
Db        558 TTAGGAGGAACAAGTC 573


RESULT 27
AV821593
LOCUS       AV821593                661 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV821593 RAFL4 Arabidopsis thaliana cDNA clone RAFL04-13-P03 5',
            mRNA sequence.
ACCESSION   AV821593
VERSION     AV821593.1  GI:19863621
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL04-13-P03"
                     /dev_stage="rosette plants"
                     /lab_host="SOLR"
                     /clone_lib="RAFL4"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     cold-treated(1,2,5,10,24 hr)"
BASE COUNT      188 a    132 c    140 g    201 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 9;  Length 661;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        139 CTTAGGAGGAACAAGT 154


RESULT 28
BJ273261/c
LOCUS       BJ273261                668 bp    mRNA    linear   EST 09-APR-2002
DEFINITION  BJ273261 Y. Ogihara unpublished cDNA library, Wh_oh Triticum
            aestivum cDNA clone whoh16g03 3', mRNA sequence.
ACCESSION   BJ273261
VERSION     BJ273261.1  GI:20098087
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whoh16g03"
                     /tissue_type="pistil at heading date"
                     /dev_stage="Feekes' scale 10.5"
                     /clone_lib="Y. Ogihara unpublished cDNA library, Wh_oh"
BASE COUNT      172 a    185 c    148 g    163 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 12;  Length 668;
```

```
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TTAGGAGGAACAAGTC 18
              ||||||||||||||||
Db        276 TTAGGAGGAACAAGTC 261


RESULT 29
BQ539794
LOCUS       BQ539794                 703 bp    mRNA    linear   EST 27-MAY-2003
DEFINITION  PTAM0102 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum
            tricornutum cDNA 5', mRNA sequence.
ACCESSION   BQ539794
VERSION     BQ539794.1  GI:21395364
KEYWORDS    EST.
SOURCE      Phaeodactylum tricornutum
  ORGANISM  Phaeodactylum tricornutum
            Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
            Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Scala,S., Carels,N., Falciatore,A., Chiusano,M.L. and Bowler,C.
  TITLE     Genome properties of the diatom Phaeodactylum tricornutum
  JOURNAL   Plant Physiol. 129 (3), 993-1002 (2002)
  MEDLINE   22111123
   PUBMED   12114555
COMMENT     Contact: Bowler C
            Laboratory of Molecular Plant Biology
            Stazione Zoologica 'Anton Dohrn'
            Villa Comunale, I-80121, Napoli, Italy
            Tel: 39 081 583 3268/3211
            Fax: 39 081 764 1355
            Email: chris@alpha.szn.it
            Seq primer: T3 backward.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Phaeodactylum tricornutum"
                     /mol_type="mRNA"
                     /db_xref="taxon:2850"
                     /cell_line="CCMP632"
                     /clone_lib="Phaeodactylum tricornutum Uni-Zap XR"
                     /note="Vector: Uni-Zap XR vector; Site_1: Eco RI; Site_2:
                     Xho I"
BASE COUNT      183 a    210 c    182 g    128 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 13;  Length 703;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 AGGAGGAACAAGTCCC 20
              ||||||||||||||||
Db        359 AGGAGGAACAAGTCCC 374


RESULT 30
BU635192
LOCUS       BU635192                 740 bp    mRNA    linear   EST 23-SEP-2002
DEFINITION  003D06 Infected Arabidopsis Leaf Arabidopsis thaliana cDNA, mRNA
            sequence.
ACCESSION   BU635192
VERSION     BU635192.1  GI:23302447
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Lundsgaard,M., Emmersen,J., Nielsen,K.L., Wilson,I., Somerville,S.
            and Welinder,K.G.
  TITLE     EST sequencing of Erysiphe cichoracearum infected Arabidopsis
            plants
  JOURNAL   Unpublished
COMMENT     Contact: Karen G. Welinder
            Institut for bioteknologi
            Aalborg Universitet
            Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
            Tel: +45 96358467
            Fax: +45 98141808
            Email: kgw@bio.auc.dk.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /dev_stage="Plant 3 weeks old, three days post infection"
                     /clone_lib="Infected Arabidopsis Leaf"
                     /note="Organ: Leaf; Vector: pBluescript; Mixed cDNA
                     library of Arabidopsis and E. cichoracearum infected leaf
                     from three weeks old Arabidopsis plants. Plants were
                     harvested 3 days after infection and mRNA oligo dT
                     selected."
BASE COUNT      205 a    142 c    170 g    223 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 13;  Length 740;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        130 CTTAGGAGGAACAAGT 145


RESULT 31
BI115344
LOCUS       BI115344                 745 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602863159F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5022326 5',
            mRNA sequence.
ACCESSION   BI115344
VERSION     BI115344.1  GI:14566245
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1842  row: e  column: 15
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5022326"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
```

for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 151 a 244 c 204 g 146 t
ORIGIN

Query Match 80.0%; Score 16; DB 12; Length 745;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 689 CTTAGGAGGAACAAGT 704

RESULT 32
BE869673
LOCUS BE869673 871 bp mRNA linear EST 20-OCT-2000
DEFINITION 601445775F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849618 5', mRNA sequence.
ACCESSION BE869673
VERSION BE869673.1 GI:10318358
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9567 row: f column: 19
High quality sequence stop: 676.
FEATURES Location/Qualifiers
source 1. .871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3849618"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_65"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies. "
BASE COUNT 274 a 169 c 161 g 267 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 871;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 180 CTTAGGAGGAACAAGT 195

RESULT 33
BI088175
LOCUS BI088175 872 bp mRNA linear EST 20-JUN-2001
DEFINITION 602851213F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4992843 5', mRNA sequence.
ACCESSION BI088175
VERSION BI088175.1 GI:14506505
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11012 row: i column: 04
High quality sequence stop: 738.
FEATURES Location/Qualifiers
source 1. .872
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4992843"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC_10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."
BASE COUNT 281 a 165 c 151 g 275 t
ORIGIN

Query Match 80.0%; Score 16; DB 12; Length 872;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTTAGGAGGAACAAGT 17
||||||||||||||||
Db 130 CTTAGGAGGAACAAGT 145

RESULT 34
BU177008
LOCUS BU177008 879 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7940818 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6155256 5', mRNA sequence.
ACCESSION BU177008
VERSION BU177008.1 GI:22690992
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13497 row: k column: 01

```
           High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6155256"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      280 a    159 c    143 g    297 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 13;  Length 879;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        439 CTTAGGAGGAACAAGT 454


RESULT 35
BG740602
LOCUS       BG740602                884 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602631028F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776247 5',
            mRNA sequence.
ACCESSION   BG740602
VERSION     BG740602.1  GI:14051255
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10628  row: h  column: 08
            High quality sequence stop: 753.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4776247"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      264 a    180 c    154 g    286 t
ORIGIN

  Query Match             80.0%;  Score 16;  DB 10;  Length 884;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 CTTAGGAGGAACAAGT 17
              ||||||||||||||||
Db        193 CTTAGGAGGAACAAGT 208


Search completed: August 15, 2003, 10:57:48
Job time : 1261 secs
```